Lecciones de física de Feynman,

Mecánica, radiación y calor

Richard P. Feynman, Robert B. Leighton
y Matthew Sands





EDICIONES
CIENTÍFICAS
UNIVERSITARIAS

TEXTO CIENTÍFICO
UNIVERSITARIO

Richard P. Feynman (Nueva York, 1918-Los Ángeles, 1988) reconocido físico estadunidense, ciencias en el Instituto Tecnológico de Massachusetts (MIT) y el doctorado en la Universidad de Princeton. Enseñó en Cornell y en el Instituto de Tecnología de California (Caltech). En 1965 recibió el Premio Nobel de Física, junto con Sin-Itiro Tomonaga y Julian Schwingerque por su trabajo fundamental en electrodinámica cuántica, con profundas consecuencias para la física de las partículas elementales.

Robert B. Leighton (Detroit, 1919-Pasadena, 1997) realizó un trabajo pionero en física de estado sólido, física de rayos cósmicos, los comienzos de la física moderna de partículas, física solar, fotografía planetaria, astronomía infrarroja y astronomía de ondas milimétricas y submilimétricas a lo largo de su vida. Fue ampliamente conocido por su diseño innovador de instrumentos científicos, y muy admirado como profesor.

Matthew Sands (Oxford, Massachusetts 1919- Santa Cruz, California, 2014), reconocido físico y profesor, doctor por el MIT. Después de la guerra, Sands ayudó a fundar la Federación de Científicos Atómicos de Los Álamos, que presionó contra el uso de armas nucleares. Participó en la construcción del sincrotrón de Caltech, fue director del Acelerador Lineal de Stanford, entre otros cargos científicos y educativos.

EDICIONES CIENTÍFICAS UNIVERSITARIAS

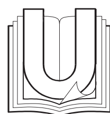
Serie Texto Científico Universitario

LECCIONES DE FÍSICA DE FEYNMAN, I

RICHARD P. FEYNMAN
ROBERT B. LEIGHTON • MATTHEW SANDS

Lecciones de física de Feynman

I. Mecánica, radiación y calor





FONDO DE CULTURA ECONÓMICA

Primera edición en inglés, 1963
Primera edición en español, 1971
Primera edición FCE, 2018
Primera edición electrónica (PDF), 2018

Feynman, Richard P., Robert B. Leighton y Matthew Sands
Lecciones de física de Feynman I. Mecánica, radiación y calor / Richard P. Feynman, Robert B. Leighton, Matthew Sands ; trad. de Enrique Oelker L., Hugo Espinosa D., Carlos Alberto Heras, Juan Martín y Marfil ; colab. de Ricardo Gómez. — México : FCE, 2018
xvii, 7, [596] p. ; ilus. ; 28 × 21 cm — (Colec. Ediciones Científicas Universitarias)
Título original: The Feynman Lectures on Physics, I. Mainly Mechanics, Radiation, and Heat
ISBN 978-607-16-5972-9 (Obra completa)
ISBN 978-607-16-5973-6 (Vol. 1)

1. Mecánica 2. Radiación 3. Calor 4. Física I. Leighton, Robert B., coaut. II. Sands, Matthew, coaut. III. Oelker L., Enrique, tr. IV. Espinosa D., Hugo, tr. V. Heras, Carlos Alberto, tr. VI. Martín y Marfil, Juan, tr. VII. Gómez, Ricardo, colab. VIII. Ser. IX. t.

LC QC23.F47 Dewey 530 F574l

Diseño de portada: Laura Esponda Aguilar
Fotografía de portada: Richard Feynman (1918-1988), físico teórico estadunidense.
Crédito: US Department of Energy/Science Photo Library

La presente edición se basa en la *New Millennium Edition* publicada por Basic Books e incluye la corrección de erratas hasta su 7ª reimpresión. Ésta es una versión corregida y revisada de la traducción de Enrique Oelker L., Hugo Espinosa D., Carlos Alberto Heras, Juan Martín y Marfil, con la colaboración de Ricardo Gómez, que fue publicada originalmente bajo el sello del Fondo Educativo Interamericano, S. A.

Título original: *The Feynman Lectures on Physics,*
I. Mainly Mechanics, Radiation, and Heat

Proyecto apoyado por el Consejo Nacional de Ciencia y Tecnología



Carretera Picacho-Ajusco, 227; 14738 Ciudad de México
www.fondodeculturaeconomica.com
Comentarios: editorial@fondodeculturaeconomica.com
Tel. (55) 5227-4672

ISBN 978-607-16-5972-9 (Obra completa)
ISBN 978-607-16-5973-6 (Vol. 1)
ISBN 978-607-16-6121-0 (pdf)
ISBN 978-607-16-6120-3 (obra completa en pdf)

Hecho en México - *Made in Mexico*

Sobre Richard Feynman

Nacido en 1918 en la ciudad de Nueva York, Richard P. Feynman recibió su doctorado de la Universidad de Princeton en 1942. A pesar de que era muy joven, tuvo un papel importante en el Proyecto Manhattan en Los Álamos durante la segunda Guerra Mundial. Más tarde, dio clases en la Universidad Cornell y en el Instituto de Tecnología de California. En 1965 recibió el premio Nobel de física, junto con Sin-Itiro Tomonaga y Julian Schwinger, por su trabajo en la electrodinámica cuántica.

El Dr. Feyman ganó el premio Nobel por haber resuelto con éxito algunos problemas de la teoría de la electrodinámica cuántica. También creó una teoría matemática que explica el fenómeno de la superfluidez en el helio líquido. Después, con Murray Gell-Mann, realizó un trabajo fundamental en el área de las interacciones débiles tales como la desintegración beta. En años posteriores, Feynman desempeñó un papel clave en el desarrollo de la teoría de los quarks al presentar su modelo de partones de procesos de colisión de protones de alta energía.

Más allá de estos logros, el Dr. Feynman introdujo nuevas técnicas básicas de computación y notación en la física: se destacan los omnipresentes diagramas de Feynman que, quizá más que ningún otro formalismo en la historia científica cercana, han cambiado la manera en la que los procesos físicos básicos son conceptualizados y calculados.

Feynman fue un maestro sobresaliente en su labor. De todos sus numerosos galardones, estaba orgulloso en particular de la Medalla Oersted para la enseñanza, a la cual se hizo acreedor en 1972. Un reseñista de *Scientific American* describió las *Lecciones de Feynman sobre física*, publicadas originalmente en 1963, como "arduas, pero sustanciosas y llenas de sabor. Después de 25 años son *la* guía para maestros y para los mejores alumnos que comienzan". Con el fin de mejorar la comprensión de la física entre los no especialistas, el Dr. Feynman escribió: *El carácter de la ley física y Electrodinámica cuántica. La extraña teoría de la luz y la materia.* También fue autor de varias publicaciones avanzadas que se han vuelto referencias y manuales clásicos para investigadores y estudiantes.

Richard Feynman fue un hombre público productivo. Su trabajo en la comisión Challenger es bien conocido, en especial su famosa demostración de la susceptibilidad de las juntas tóricas al frío, un experimento elegante que requería de nada más que un vaso de agua helada y una abrazadera. Fueron menos célebres los esfuerzos del Dr. Feynman en los años sesenta en el comité curricular de la Universidad Estatal de California, donde protestó por la mediocridad de los libros de texto.

Una recitación de los incontables logros científicos y educativos de Richard Feynman no puede capturar adecuadamente la esencia de este hombre. Como lo sabe cualquier lector de sus publicaciones, incluso las más técnicas, la personalidad dinámica y multifacética de Feynman resplandece a lo largo de toda su obra. Además de físico, en muchas ocasiones fue reparador de radios, forzador de cerraduras, artista, bailarín, intérprete de bongó y descifrador de jeroglíficos mayas. Con una curiosidad perpetua por su mundo, fue un empirista ejemplar.

Richard Feynman murió el 15 de febrero de 1988 en Los Ángeles.

Prefacio a la Edición del nuevo milenio

Han pasado casi 50 años desde que Richard Feynman impartió el curso introductorio de física en el Instituto de Tecnología de California (Caltech) que dio origen a estos tres volúmenes, las *Lecciones de Feynman sobre física.* En esos 50 años, nuestra comprensión del mundo físico ha cambiado en gran medida, pero esta obra ha perdurado. Las lecciones de Feynman son tan poderosas hoy en día como lo eran recién publicadas, gracias a lo singular de sus observaciones sobre la física y su método de enseñanza. Físicos principiantes y experimentados las han estudiado por igual en todo el mundo; han sido traducidas al menos a una docena de lenguas, con más de 1.5 millones de ejemplares impresos sólo en inglés. Quizá ninguna otra serie de libros de física haya tenido un impacto tan amplio y duradero.

Esta *Edición del nuevo milenio* abre paso a una nueva era para las *Lecciones de Feynman* sobre física: la era de la edición digital en el siglo XXI. Las *lecciones* se han vuelto *Lecciones-e* cuyo texto y ecuaciones se expresan en el lenguaje de composición tipográfica LaTeX y cuyas figuras fueron hechas de nuevo en su totalidad con software de dibujo moderno.

Las consecuencias para la edición *impresa* de esta edición *no* sorprenden; se ve casi igual a los libros rojos originales que los estudiantes de física han conocido y disfrutado por décadas. Las diferencias principales están en un índice alfabético ampliado y mejorado, en la corrección de 885 erratas que los lectores encontraron a lo largo de cinco años desde la primera impresión de la edición anterior, y en la facilidad de corregir erratas que los futuros lectores lleguen a encontrar. Regresaré a esto líneas adelante.

La *Versión electrónica* de esta edición y la *Versión electrónica aumentada* son innovaciones electrónicas. En contraste con la mayoría de las versiones electrónicas de libros técnicos del siglo XX, cuyas ecuaciones, figuras y en ocasiones incluso texto pierden calidad en cuanto uno trata de ampliarlos, el manuscrito en LaTeX de la *Edición del nuevo milenio* hace posible la creación de libros electrónicos de la más alta calidad, en los que todas las características de la página (excepto las fotografías) pueden ampliarse sin límites y conservar sus formas y definición precisas. Y la *Versión electrónica aumentada*, que cuenta con audio y fotografías de pizarrones de las lecciones originales de Feynman, además de hipervínculos a otros recursos, es una innovación que le habría dado mucha satisfacción a Feynman.

Recuerdos de las lecciones de Feynman

Estos tres volúmenes son un tratado pedagógico autosuficiente. También son un registro histórico de las lecciones universitarias de física de Feynman entre 1961 y 1964, una materia obligatoria para todos los estudiantes de primer y segundo año en Caltech, sin importar la especialidad a la que aspirarán.

Los lectores se preguntarán, como lo hice yo, cuál fue el impacto que las lecciones de Feynman tuvieron en los estudiantes que las presenciaron. En su Prefacio a estos volúmenes, Feynman ofreció un punto de vista algo negativo al respecto: “No creo que me haya ido muy bien con los estudiantes”, escribió. Matthew Sands, en su breve texto autobiográfico en *Feynman's Tips on Physics*, expresó un punto de vista mucho más optimista. Por pura curiosidad, en la primavera de 2005 envié correos electrónicos o hablé con una muestra cuasi aleatoria de 17 alumnos (de unos 150) de la clase de Feynman entre 1961 y

1963 —entre ellos algunos a los que se les dificultó mucho la clase y otros que no tuvieron ningún problema para dominarla, biólogos, químicos, ingenieros, geólogos, matemáticos y astrónomos, al igual que físicos—.

Puede que los años intermedios hayan esmaltado sus recuerdos de un matiz eufórico, pero alrededor del 80 % de ellos recuerdan las lecciones de Feynman como uno de los puntos más memorables de sus años de universidad. "Era como ir a la iglesia". Las lecciones eran "una experiencia transformativa", "una experiencia de vida, probablemente lo más importante que obtuve de Caltech". "Yo estudiaba biología, pero las lecciones de Feynman se destacan como uno de los puntos más altos en mi experiencia universitaria... aunque debo admitir que en esa época no podía hacer la tarea y casi no entregué ninguna". "Yo me encontraba entre los estudiantes menos prometedores de su clase, y nunca me perdí una sola sesión... recuerdo y todavía puedo sentir la dicha de Feynman por el descubrimiento... Sus lecciones tenían un... impacto emocional que quizá se perdió en las *Lecciones* impresas".

En contraste, varios de los estudiantes tienen recuerdos negativos debido en gran medida a dos cuestiones: (i) "No podías aprender a trabajar con los problemas de tarea al asistir a las lecciones. Feynman era demasiado ingenioso: conocía trucos y sabía qué aproximaciones podían hacerse, tenía una intuición basada en la experiencia y en una genialidad que un estudiante que apenas comienza no posee". Feynman y sus colegas, conscientes de esta falla en el curso, la resarcieron en parte con materiales que se han incorporado a *Feynman's Tips on Physics*: tres lecciones hechas por el propio Feynman sobre resolución de problemas, y un conjunto de ejercicios y respuestas preparado por Robert B. Leighton y Rochus Vogt. (ii) "La inseguridad de no saber lo que podría discutirse en la siguiente lección, la falta de un libro de texto o referencia que tuviera cualquier relación con el material de lectura y la consecuente inhabilidad que teníamos para adelantar lecturas eran muy frustrantes... Las lecciones me parecían emocionantes y comprensibles en el aula, pero afuera eran sánscrito (cuando trataba de reconstruir los detalles)". Desde luego, este problema se resolvió en estos tres volúmenes, la versión impresa de las *Lecciones de Feynman sobre física*. Éstas se volvieron el libro de texto que los estudiantes de Caltech han usado para estudiar durante muchos años después, y sobreviven hoy en día como uno de los mayores legados de Feynman.

Una historia de erratas

Feynman y sus coautores, Robert B. Leighton y Matthew Sands, produjeron las *Lecciones de Feynman sobre física* con gran rapidez, trabajando a partir de grabaciones y fotografías de pizarrones de las lecciones de Feynman* (ambas se incorporaron en la *Versión electrónica aumentada* de esta *Edición del nuevo milenio*). Debido al ritmo tan acelerado al que trabajaron Feynman, Leighton y Sands, era inevitable que muchos errores se colaran en la primera edición. Feynman acumuló largas listas de erratas alegadas en los años subsecuentes, erratas que hallaron los estudiantes y profesores de Caltech así como lectores alrededor del mundo. En los años sesenta e inicios de los setenta, Feynman hizo un espacio en su ajetreada vida para verificar la mayor parte de las erratas, aunque no todas, de los volúmenes I y II e insertar las correcciones en las reimpresiones siguientes. Pero el sentido del deber de Feynman nunca logró sobreponerse a su entusiasmo por nuevos descubrimientos para lograr que se ocupara de las erratas en el volumen III†. Después de su muerte prematura en 1988, las listas de erratas de los tres volúmenes se guardaron en el Archivo de Caltech y quedaron ahí olvidadas.

---

* Para descripciones de la génesis de las lecciones de Feynman y de estos volúmenes, véanse el Prefacio y las Introducciones a cada uno de los tres volúmenes, y también el texto autobiográfico de Matt Sands en *Feynman's Tips on Physics*, y el Prefacio especial a la *Edición conmemorativa* de las *Lecciones de Feynman*, escrito en 1989 por David Goodstein y Gerry Neugebauer, que también aparece en la *Edición definitiva* de 2005.

† En 1975 comenzó a buscar las erratas del volumen III, pero se distrajo con otras cosas y nunca terminó la tarea, de modo que no se hicieron correcciones.

En 2002, Ralph Leighton (hijo del fallecido Robert Leighton y compañero de Feynman) me informó de las viejas erratas y de una nueva y larga lista que había compilado su amigo Michael Gottlieb. Leighton propuso que Caltech realizara una nueva edición de las *Lecciones de Feynman* con todas las erratas corregidas, y que la publicara junto con un nuevo volumen de material complementario, *Feynman's Tips on Physics*, que él y Gottlieb estaban preparando.

Feynman era mi héroe y fue mi amigo cercano. Cuando vi las listas de erratas y el contenido del nuevo volumen propuesto, accedí pronto a supervisar el proyecto en representación de Caltech (que durante mucho tiempo fue el hogar académico de Feynman, al cual él, Leighton y Sands habían encomendado todos los derechos y responsabilidades de las *Lecciones de Feynman*). Después de un año y medio de trabajo meticuloso por parte de Gottlieb, y del atento examen del Dr. Michael Hartl (un investigador posdoctoral notable de Caltech que evaluó todas las erratas además del nuevo volumen), la *Edición definitiva de las lecciones de Feynman sobre física* vio la luz en 2005; tenía alrededor de 200 erratas corregidas y estaba acompañada por *Feynman's Tips on Physics*, de Feynman, Gottlieb y Leighton.

Yo *pensé* que esa edición iba a ser la "definitiva". Lo que no anticipé fue la respuesta entusiasta de los lectores alrededor del mundo ante el llamado de Gottlieb para identificar más erratas y enviarlas por medio del sitio web que él creó y aún mantiene, *The Feynman Lectures Website*, `www.feynmanlectures.info`. En los cinco años a partir de entonces, se han enviado 965 erratas nuevas que sobrevivieron el meticuloso escrutinio de Gottlieb, Hartl y Nate Bode (un estudiante de posgrado notable de Caltech, quien sucedió a Hartl como el examinador de erratas por parte de esta institución). De estas 965 erratas evaluadas, 80 se corrigieron en la tercera reimpresión de la *Edición definitiva* (agosto de 2006) y las otras 885 se corrigieron en esta primera impresión de la *Edición del nuevo milenio* (332 en el volumen I, 263 en el volumen II y 200 en el volumen III). Para los detalles de las erratas, véase `www.feynmanlectures.info`.

Está claro que liberar de errores las *Lecciones de Feynman sobre física* se ha vuelto una empresa comunitaria a nivel mundial. En nombre de Caltech agradezco a los 50 lectores que han contribuido desde 2005 y a los muchos más que lleguen a contribuir en los años por venir. Los nombres de todos los que contribuyeron se encuentran en `www.feynmanlectures.info/flp_errata.html`.

Casi todas las erratas han sido de tres tipos: (i) errores tipográficos en prosa; (ii) errores tipográficos y matemáticos en ecuaciones, tablas y figuras —errores en signos, números incorrectos (por ejemplo, un 5 que debería ser un 4), así como subíndices, signos de suma, paréntesis y términos faltantes en ecuaciones—; (iii) referencias cruzadas incorrectas a otros capítulos, tablas y figuras. Esta clase de errores, a pesar de que no son gravísimos para un físico experimentado, pueden ser frustrantes y confusos para la audiencia primaria de Feynman: los estudiantes.

Cabe resaltar que entre las 1165 erratas que se corrigieron bajo mi auspicio son pocas las que considero como verdaderos errores en física. Un ejemplo está en el volumen II, página 5-10 que ahora dice "[...] si una cavidad está completamente encerrada por un conductor a tierra, ninguna distribución estática de cargas en el *interior* puede producir campos [eléctricos] en el *exterior*" (la precisión "a tierra" fue omitida en ediciones previas). Este error le fue señalado a Feynman por varios lectores, entre los que estaba Beulah Elizabeth Cox, estudiante en el College of William and Mary, quien confió en el pasaje erróneo de Feynman para un examen. En 1975 Feynman le escribió lo siguiente a la señorita Cox:* "Tu profesor tuvo razón en no darte puntos, pues tu respuesta es incorrecta, como lo demostró él usando la ley de Gauss. En la ciencia, se deben usar argumentos lógicos y elaborados con cuidado, no sólo creer en las autoridades. También leíste el libro correctamente y lo entendiste. Cometí un error, y el libro está mal. Tal vez estaba pensando en una esfera conductora a tierra, o bien en el hecho de que mover las cargas en lugares distintos del interior no afecta las cosas en el exterior. No estoy seguro, pero metí la pata. Y tú también metiste la pata por creerme."

---

* Páginas 288-289 de *Perfectly Reasonable Deviations from the Beaten Track, The Letters of Richard P. Feynman*, editado por Michelle Feynman (Basic Books, Nueva York, 2005).

Cómo cobró vida esta Edición del nuevo milenio

Entre noviembre de 2005 y julio de 2006, 340 erratas fueron enviadas al sitio web de las *Lecciones de Feynman* `www.feynmanlectures.info`. Llama la atención que el grueso de éstas viniera de una sola persona: el Dr. Rudolf Pfeiffer, que entonces era becario de posdoctorado en la Universidad de Viena, Austria. La editorial, Addison Wesley, arregló 80 erratas, pero se resistió a arreglar más debido al costo: los libros se imprimían mediante un proceso de foto offset y trabajaban con imágenes fotográficas de las páginas de los años sesenta. Corregir un error implicaba volver a formar la página entera, y para asegurar que no se colaran nuevos errores, la página debía volver a formarse por dos personas diferentes, para que luego la leyeran y corrigieran varias personas distintas: se trata de un proceso muy costoso cuando hay cientos de erratas en juego.

Gottlieb, Pfeiffer y Ralph Leighton estaban muy disgustados con esto, por lo que formularon un plan para facilitar el arreglo de todas las erratas, y también producir la versión electrónica y la aumentada de las *Lecciones de Feynman sobre física*. Me propusieron su plan a mí, como representante de Caltech, en 2007. Me emocioné, pero procedí con cautela. Después de ver más detalles, entre ellos un capítulo de muestra de la *Versión electrónica aumentada*, recomendé que Caltech cooperara con Gottlieb, Pfeiffer y Leighton para llevar a cabo su plan. El plan recibió la aprobación de tres catedráticos sucesivos de la División de Física, Matemáticas y Astronomía de Caltech: Tom Tombrello, Andrew Lange y Tom Soifer, y los detalles contractuales y legales complejos fueron resueltos por el abogado de Propiedad Intelectual de Caltech, Adam Cochran. Con la publicación de esta *Edición del nuevo milenio*, el plan se llevó a cabo con éxito, a pesar de su complejidad. Específicamente:

Pfeiffer y Gottlieb convirtieron a LaTeX los tres volúmenes de las *Lecciones* (y también más de 1000 ejercicios del curso de Feynman para incorporarlos a *Feynman's Tips on Physics*). Las figuras de las *Lecciones* se volvieron a trazar en formato electrónico moderno en India, bajo la supervisión del traductor al alemán de las *Lecciones*, Henning Heinze, para su uso en la edición alemana. Gottlieb y Pfeiffer intercambiaron el uso no exclusivo de sus ecuaciones en LaTeX en la edición alemana (publicada por Oldenbourg) por el uso no exclusivo de las figuras de Heinze en esta *Edición del nuevo milenio* inglesa. Pfeiffer y Gottlieb verificaron con cuidado todo el texto y ecuaciones en LaTeX y todas las figuras que se volvieron a trazar, e hicieron las correcciones necesarias. Nate Bode y yo, en representación de Caltech, hemos hecho controles aleatorios del texto, ecuaciones y figuras; y sorprendentemente, no hemos encontrado errores. Pfeiffer y Gottlieb son meticulosos y precisos en un grado increíble. Gottlieb y Pfeiffer concertaron que John Sullivan de la Biblioteca Huntington digitalizara las fotos de los pizarrones de Feynman entre 1962 y 1964, y que George Blood Audio digitalizara las grabaciones de las lecciones —con apoyo financiero y moral del profesor de Caltech Carver Mead, apoyo logístico de la archivista Shelley Erwin, y apoyo legal de Cochran—.

La problemática legal era seria: en los años sesenta, Caltech cedió los derechos a Addison Wesley para publicar la edición impresa, y en los años noventa, los derechos para distribuir el audio de las lecciones de Feynman y una variante de una edición electrónica. En la década del 2000, a través de una serie de adquisiciones de esas licencias, los derechos impresos se transfirieron al grupo editorial Pearson, mientras que los derechos electrónicos y de audio se transfirieron al grupo editorial Perseus. Con ayuda de Ike Williams, abogado especialista en lo editorial, Cochran logró unir todos estos derechos con Perseus (Basic Books), con lo que hizo posible esta *Edición del nuevo milenio*.

Agradecimientos

En nombre de Caltech, agradezco a las numerosas personas que han hecho posible esta *Edición del nuevo milenio*. En específico, quiero agradecer a las personas clave mencionadas líneas arriba: Ralph Leighton, Michael Gottlieb, Tom Tombrello, Michael Hartl, Rudolf Pfeiffer, Henning Heinze, Adam Cochran, Carver Mead, Nate Bode, Shelley Erwin, Andrew Lange, Tom Soifer, Ike Williams, y a las 50 personas que enviaron las erratas (quienes se enlistan en `www.feynmanlectures.info`). También agradezco a Michelle Feynman (hija de Richard Feynman) por su apoyo y consejo continuos, a Alan Rice por su asistencia tras bambalinas y asesoría en Caltech, a Stephan Puchegger y Calvin Jackson por su ayuda y orientación, a Pfeiffer por la conversión de las *Lecciones* a LaTeX, y a Michael Figl, Manfred Smolik y Andreas Stangl por las discusiones sobre correcciones de erratas; así como al personal de Perseus/Basic Books, y (por las ediciones previas) al personal de Addison Wesley.

Kip S. Thorne
Profesor emérito Feynman de física teórica
Instituto de Tecnología de California *octubre de 2010*



Prefacio de Feynman

Estas son las lecciones de física que di el año pasado y el antepasado a los estudiantes de los primeros años en el Instituto Tecnológico de California (Caltech). Por supuesto que éstas no son textuales; han sido editadas, a veces con gran extensión y a veces con muy poca. Las clases constituyen sólo parte del curso completo. El grupo total de 180 estudiantes se reunía dos veces por semana en un aula grande para atender a las explicaciones, y luego se dividía en grupos pequeños de 15 a 20 estudiantes en secciones de discusión y práctica bajo la guía de un ayudante de docencia. Había, además, una sección de laboratorio por semana.

Con estas clases tratábamos de resolver un problema especial: mantener el interés de los estudiantes muy entusiastas y bastante despiertos que regresaban de la escuela secundaria para entrar en el Caltech. Muchos habían oído hablar de lo interesante y estimulante que es la física: la teoría de la relatividad, la mecánica cuántica y otras ideas modernas. Pero al terminar los dos años del curso anterior al nuestro, muchos de ellos se sentían descorazonados porque realmente se les presentaban muy pocas ideas geniales, nuevas o interesantes. Se les hacía estudiar planos inclinados, electrostática y cuestiones por el estilo, y después de dos años era como para volverse tonto. Se trataba, pues de ver si podíamos hacer un curso que salvara a los estudiantes más avanzados y animados manteniéndoles el entusiasmo.

Aunque mi intención no fue convertir las clases en un campo de estudio e investigación preparé las lecciones para los más inteligentes de la clase a fin de asegurarme, si era posible, que aun los estudiantes más inteligentes no podrían abarcar completamente el contenido de cada lección; para ello introduje recomendaciones sobre la aplicación de las ideas y conceptos en diversas direcciones, las cuales se apartaban de la línea principal de ataque. Por esta razón, sin embargo, traté concienciudamente de hacer que todos los enunciados fueran lo más precisos posibles de señalar en cada caso dónde encajaban las ecuaciones en el cuerpo de la física y cómo —cuando aprendieron más— se modificarían las cosas. Pensé, además, que para los estudiantes es importante indicar qué es lo que deben ser capaces —si son suficientemente inteligentes— de comprender por deducción de lo que se ha dicho antes y qué se está introduciendo como cosa nueva. Cuando se presentaban nuevas ideas, traté de deducirlas si eran deducibles o de explicar que era una nueva idea sin base alguna en lo que ya habían aprendido y que se suponía que no era demostrable, sino simplemente un agregado.

Al comenzar estas lecciones supuse que los estudiantes traían cierto conocimiento de la escuela secundaria —tal como óptica geométrica, ideas simples sobre química, etcétera—. Creí, también, que no había ninguna razón para dictar las lecciones en un orden determinado, en el sentido de que no debería mencionar algo hasta que no estuviera en condiciones de estudiarlo en detalle. Había abundante material que surgiría, pero sin discusión completa. Las discusiones completas se harían más tarde cuando hubiera mayor preparación para seguirlas. Ejemplos de esto son el tratamiento de la inductancia y el de los niveles de energía, que primero se introducen en forma muy cualitativa y luego se desarrollan en forma más detallada.

Al mismo tiempo que me dirigía al estudiante más activo, quise cuidar del individuo para quien los adornos en demasía y las aplicaciones laterales son meramente intranquilizadores, cuando no se puede esperar, de ningún modo, que aprendan la mayor parte del material. Para ese estudiante traté que hubiera un núcleo central o columna vertebral de material que pudiera captar. Tenía la esperanza de que no se pondría nervioso aunque no entendiera todo el contenido de una lección. No esperaba que comprendiera todo sino los rasgos centrales y más directos. Naturalmente que se necesitaba cierta inteligencia de su parte para ver cuáles eran los teoremas e ideas centrales y cuáles los resultados y aplicaciones laterales más avanzados que sólo podría entender en años posteriores.

Había una dificultad seria para dar estas lecciones: por la forma en que se daba el curso, no había una retroacción del estudiante hacia el profesor que indicara cómo estaban yendo las lecciones. Esta es una dificultad muy seria que me impide saber con certeza hasta qué punto, en realidad, fueron muy provechosas mis clases. Todo era esencialmente un experimento. Y si lo hiciera nuevamente no lo haría en la misma forma —¡espero que no tenga que hacerlo de nuevo!— Creo, no obstante, que en lo que respecta a la física las cosas anduvieron muy satisfactoriamente el primer año.

El segundo año no quedé tan satisfecho. En la primera parte del curso, que trataba de la electricidad y el magnetismo, no pude encontrar ningún modo realmente único o diferente de hacerlo —ninguna manera que fuera especialmente más estimulante que la forma habitual de presentarlo—. Por lo tanto, no creo que hice mucho en las clases sobre electricidad y magnetismo. Al final del segundo año había pensado originalmente continuar dando, después de la electricidad y el magnetismo, algunas clases sobre las propiedades de los materiales, pero con el interés de explicar los modos fundamentales, las soluciones de la ecuación de difusión, los sistemas vibrantes, las funciones ortogonales... desarrollando, así, las primeras etapas de lo que usualmente se denomina "métodos matemáticos de la fisica". Pensando en retrospectiva, creo que si lo hiciera de nuevo volvería a esa idea original. Pero como no se había contemplado que yo daría estas clases nuevamente, se sugirió que sería conveniente tratar de dar una introducción a la mecánica cuántica —que es lo que ustedes encontrarán en el tercer volumen—.

Queda perfectamente claro que los estudiantes que sigan estudiando física deberán esperar hasta el tercer año para estudiar mecánica cuántica. Por otra parte, se esgrimió el argumento de que muchos de los estudiantes de nuestro curso estudian física como base para su especialización en otros campos. Y la forma habitual de tratar la mecánica cuántica hace que el tema sea casi inalcanzable para la gran mayoría de los estudiantes debido a que necesitan mucho tiempo para aprenderlo. No obstante, en sus aplicaciones concretas, —especialmente en sus aplicaciones más complejas, como en la ingeniería eléctrica y en la química— realmente no se usa la maquinaria completa del tratamiento con ecuaciones diferenciales. Por ello, traté de describir los principios de la mecánica cuántica de un modo que no exigiera un conocimiento básico de la matemática de las ecuaciones diferenciales. Creo que aun para un físico es muy interesante presentar la mecánica cuántica de esta manera inversa —por varias razones que se pueden ver en las lecciones mismas—. Sin embargo, creo que el experimento en lo correspondiente a la mecánica cuántica no tuvo éxito completo —en gran parte debido a que no tuve tiempo al final (por ejemplo, debería haber tenido tres o cuatro lecciones más para tratar detenidamente temas tales como las bandas de

energía y la dependencia espacial de las amplitudes)—. Además, nunca había presentado antes el tema de este modo, por lo que la falta de retroacción fue particularmente seria. Ahora creo que se debe dar la mecánica cuántica más tarde. A lo mejor algún día tenga la oportunidad de hacerlo de nuevo. Entonces lo haré mejor.

La razón de que no haya lecciones sobre cómo resolver problemas se debe a que había secciones de discusión y práctica. Aunque en el primer año incluí tres lecciones sobre cómo resolver problemas, en este curso no pude hacerlo. También había una lección sobre guía inercial que debe estar ciertamente después de la lección sobre sistemas que rotan, pero que desafortunadamente se omitió. Las lecciones quinta y sexta fueron dadas por Matthew Sands porque yo estaba fuera de la ciudad.

La pregunta es, por supuesto, hasta qué punto este experimento ha tenido éxito. Mi punto de vista —que, sin embargo, no lo comparten la mayoría de los que trabajaron con los estudiantes— es pesimista. No creo haber obtenido gran éxito en lo que respecta a ellos. Cuando recuerdo el modo en que éstos manipulaban los problemas en los exámenes, pienso que el sistema es un fracaso. Por supuesto, mis amigos me indican que hubo una o dos docenas de estudiantes que —muy sorprendentemente— comprendieron casi todo el contenido en todas las lecciones y que fueron muy activos trabajando con el material y preocupándose con animación e interés por muchos tópicos. Estos individuos tienen actualmente, creo, una base de primera línea en física —y son, después de todo, aquellos a quienes quería llegar—. Pero entonces "El poder de la instrucción es, en general, poco eficaz, excepto en las felices disposiciones en que es casi superfluo" (Gibbons).

De todos modos, no quería dejar ningún estudiante completamente atrasado, como quizás lo hice. Creo que un modo más efectivo de ayudar al estudiante sería mediante la adición de tiempo y esfuerzo en el desarrollo de un conjunto de problemas que aclare algunas de las ideas contenidas en las lecciones. Los problemas dan la oportunidad de aumentar la comprensión del material expuesto haciéndolo más real, estructurado y accesible para el proceso de fijación.

Pienso, sin embargo, que la solución a este problema educativo no es otra que darse cuenta que la enseñanza sólo puede realizarse cuando hay una relación individual directa entre un estudiante y un buen profesor, situación en la cual el estudiante discute las ideas, piensa en las cosas y habla sobre ellas. Es imposible aprender simplemente asistiendo a una clase, o simplemente resolviendo los problemas asignados. Pero en los actuales momentos tenemos tantos estudiantes a quienes enseñar que tenemos que encontrar un sustituto de lo ideal. Quizás mis lecciones ejerzan alguna contribución. Quizás en algún lugar pequeño donde sea posible una relación individual entre profesores y estudiantes, éstos obtengan alguna inspiración o algunos conceptos de estas lecciones. Quizás entonces, también el proceso de fijar el material sea más alegre y placentero para ellos y dé origen al desarrollo de algunas ideas.

Richard P. Feynman

junio 1963

Introducción

Este libro está basado en el curso de lecciones introductorias de física que dio el profesor R. P. Feynman en el Instituto de Tecnología de California (Caltech) durante el año académico 1961-1962; abarca el primero de los dos años del curso introductorio que toman todos los estudiantes de primer y segundo año en Caltech, y a éste siguió en 1962-1963 una serie similar que abarcó el segundo año. Las lecciones constituyen una parte muy importante en la revisión del curso introductorio, la cual se llevó a cabo a lo largo de un periodo de cuatro años.

La necesidad de una revisión básica surgió tanto del rápido desarrollo de la física en las décadas recientes como del hecho de que los estudiantes recién llegados han mostrado un aumento constante en su habilidad matemática a partir de las mejoras en el contenido de las clases de matemáticas de preparatoria. Nuestra intención fue aprovechar este mayor bagaje matemático y al mismo tiempo introducir suficientes temas modernos para que el curso fuera exigente, interesante y más representativo de la física de hoy en día.

Para generar una variedad de ideas sobre qué material incluir y cómo presentarlo, se alentó a un número sustancial de profesores de física a que ofrecieran sus ideas en forma de planes temáticos para un curso revisado. Varios de éstos fueron presentados y discutidos de manera crítica y profunda. Casi de inmediato se llegó al acuerdo de que una revisión básica del curso no se podría lograr adoptando un libro de texto diferente, ni siquiera escribiendo uno ab initio, sino que el nuevo curso debía centrarse en un conjunto de lecciones presentadas a ritmo de dos o tres por semana; el material textual apropiado se produciría luego como una operación secundaria conforme avanzaba el curso, y también se organizarían experimentos de laboratorio adecuados para el material de las lecciones. En consecuencia, se fijó un plan del curso a grandes rasgos, pero se reconoció que era tentativo y estaba incompleto y sujeto a modificaciones considerables por parte de quien fuera a tener la responsabilidad de efectivamente preparar las lecciones.

En cuanto al mecanismo mediante el cual sería finalmente puesto en práctica el curso, se consideraron varios planes. Éstos se parecían mucho entre sí, incluían un esfuerzo cooperativo por N número de miembros del personal que compartirían la carga total de trabajo de manera simétrica y equitativa: cada uno tomaría la responsabilidad de 1/N del material, daría las lecciones y escribiría el material textual para su parte. Sin embargo, la falta de disponibilidad de personal suficiente y la dificultad de mantener un punto de vista uniforme ante las diferencias de personalidad y filosofía entre los participantes individuales hacían que estos planes fueran irrealizables.

Darse cuenta de que verdaderamente poseíamos los medios para crear no sólo un curso nuevo y diferente de física, sino quizás uno único, resultó ser una inspiración dichosa para el profesor Sands. Él sugirió que el profesor R. P. Feynman preparara y diera las lecciones, y que éstas fueran grabadas. Una vez transcritas y editadas, se convertirían en el libro de texto para el nuevo curso. En esencia, éste fue el plan que se adoptó.

Se esperaba que la edición necesaria fuera menor, pues consistiría principalmente en proporcionar figuras y revisar la puntuación y la gramática; la harían uno o dos estudiantes de posgrado a tiempo parcial. Desafortunadamente, esta expectativa fue pasajera. En realidad, fue una operación editorial mayor poder transformar la transcripción literal a una forma legible, aun sin la reorganización o revisión del tema tratado que a veces se requería. Además, no era trabajo para un editor técnico o para un estudiante de posgrado, sino uno que requería la atención minuciosa de un físico profesional ¡de 10 a 20 horas por lección!

La dificultad de la tarea editorial, junto con la necesidad de poner el material en manos de los estudiantes tan pronto como fuera posible, puso un límite estricto en qué tanto se podía “pulir” el material en realidad, y en consecuencia nos vimos forzados a apostar por un producto preliminar pero técnicamente correcto que pudiera usarse de inmediato, en lugar de uno que pudiera considerarse final o terminado. Debido a una necesidad urgente de más ejemplares para nuestros alumnos, y a un interés alentador por parte de profesores y estudiantes de varias instituciones distintas, decidimos publicar el material en su forma preliminar en lugar de esperar una revisión ulterior más profunda que quizá nunca llegaría. No nos hacemos ilusiones en cuanto a lo completo o fluido del material, ni en cuanto a su organización lógica; de hecho, planeamos varias modificaciones menores en el futuro inmediato, y esperamos que no se vuelva estático en su forma o contenido.

Además de las lecciones, que constituyen una parte de importancia central para el curso, era necesario también proporcionar ejercicios adecuados para acrecentar la experiencia y habilidad de los estudiantes, así como experimentos oportunos para facilitar un contacto de primera mano con el material de las lecciones en el laboratorio. Ninguno de estos aspectos se encuentra en un estado tan avanzado como el material de las lecciones, pero se ha hecho un progreso considerable. Algunos ejercicios fueron elaborados conforme progresaban las lecciones y luego expandidos y amplificados para su uso al año siguiente. Sin embargo, puesto que todavía no estamos convencidos de que los ejercicios proporcionen suficiente variedad y profundidad de aplicación de los materiales de las lecciones para dar al alumno plena conciencia del tremendo poder que está a su disposición, los ejercicios se publicaron por separado en un formato menos permanente para así alentar una revisión frecuente.

El profesor H. V. Neher ha ideado varios experimentos nuevos para el nuevo curso. Entre ellos se encuentran varios que utilizan la fricción extremadamente baja que presenta un cojinete de gas: un nuevo canal lineal de aire con el que es posible hacer mediciones cuantitativas de movimiento unidimensional, impactos y movimiento armónico, y un trompo de color de Maxwell soportado e impulsado por aire con el que se pueden estudiar el movimiento de rotación acelerado y la precesión y nutación giroscópicas. Se espera que el desarrollo de nuevos experimentos de laboratorio dure un tiempo considerable.

El programa de revisión se llevó a cabo bajo la dirección de los profesores R. B. Leighton, H. V. Neher y M. Sands. En el programa participaron oficialmente los profesores R. P. Feynman, G. Neugebauer, R. M. Sutton, H. P. Stabler, F. Strong y R. Vogt, de la división de física, matemáticas y astronomía, y los profesores T. Caughey, M. Plesset y C. H. Wilts de la división de ciencias de la ingeniería. Reconocemos y agradecemos la valiosa asistencia de todos aquellos que contribuyeron al programa de revisión. Estamos en deuda en especial con la Fundación Ford, sin cuyo apoyo financiero este programa no podría haberse llevado a cabo.

Robert B. Leighton

julio 1963

Índice general

1

Átomos en movimiento

1-1 Introducción

Este curso de física de dos años se presenta desde el punto de vista de que usted, el lector, va a ser un físico. Este, desde luego, no necesariamente es el caso, pero, ¡es lo que supone cada profesor en cada tema! Si usted va a ser un físico, tendrá que estudiar mucho: 200 años del campo de conocimiento de más rápido desarrollo que existe. Tanto conocimiento, en efecto, que usted pensaría que no lo podría aprender todo en cuatro años y realmente no puede; ¡deberá ir, además, a la escuela de graduados!

Es bastante sorprendente que, a pesar de la cantidad tremenda de trabajo que ha sido realizado en todo este tiempo, es posible condensar en gran parte la enorme masa de resultados (esto es, hallar *leyes* que resumen todo nuestro conocimiento). Aun así, las leyes son tan difíciles de entender que es injusto para usted comenzar a explorar este tremendo tema sin algún tipo de mapa o bosquejo de la relación de una parte a la otra del contenido de la ciencia. De acuerdo con estas notas preliminares, los primeros tres capítulos darán, por lo tanto, un bosquejo de la relación de la física con el resto de las ciencias, las relaciones de las ciencias entre sí y el significado de ciencia, para ayudarnos a desarrollar una mejor comprensión de este tema.

Usted podría preguntar por qué no podemos enseñar física dando simplemente las leyes básicas en la página uno y luego mostrar cómo operan en todas las circunstancias posibles, como lo hacemos en geometría euclidiana, donde establecemos los axiomas y luego hacemos toda clase de deducciones. (Así que, no satisfecho con aprender física en cuatro años, ¿la quiere aprender en sólo cuatro minutos?) No lo podemos hacer de esta manera por dos razones. Primero, aún no *conocemos* todas las leyes básicas: existe una frontera de ignorancia en expansión. Segundo, el planteamiento correcto de las leyes de la física contiene algunas ideas muy poco familiares que requieren matemática avanzada para su descripción. Por esta razón se necesita una cantidad apreciable de entrenamiento preparatorio para entender lo que significan estas *palabras.* No, no es posible hacerlo de esta manera. Sólo podemos hacerlo parte por parte.

Cada parte del todo de la naturaleza es siempre sólo una *aproximación* a la verdad completa o la verdad completa hasta donde la conocemos. En realidad, todo lo que sabemos es sólo una cierta forma de aproximación, porque *sabemos que aún no conocemos todas las leyes.* Por eso, las cosas deben ser aprendidas, sólo para luego desaprenderlas o, más a menudo, para corregirlas.

El principio de la ciencia, casi la definición, es la siguiente: *La comprobación de todo conocimiento es el experimento.* El experimento es el *único juez* de la "verdad" científica. Pero ¿cuál es la fuente del conocimiento? ¿De dónde provienen las leyes que deben ser comprobadas? El experimento mismo ayuda para producir estas leyes, en el sentido de que nos da sugerencias. Pero también la *imaginación* es necesaria para crear, basada en estas sugerencias, las grandes generalizaciones (adivinar sus admirables, simples, pero muy extraños esquemas que hay detrás de todas ellas, y luego experimentar para comprobar nuevamente si hemos hecho la suposición correcta). Este proceso de imaginación es tan difícil que existe una división de las labores en física: hay físicos *teóricos* que imaginan, deducen y hacen suposiciones acerca de nuevas leyes, pero no experimentan; y luego hay los físicos *experimentales* que experimentan, imaginan, deducen y hacen suposiciones.

Decíamos que las leyes de la naturaleza son aproximadas: que primero encontramos las "equivocadas" y luego encontramos las "correctas". Pero ¿cómo puede un experimento estar "equivocado"? Primero, en forma trivial: si algo anda mal con los aparatos y ustedes no se dieron cuenta. Pero estas cosas se arreglan fácilmente y se comprueban en todos los sentidos. Pero sin aferrarnos a estos problemas menores, ¿cómo *pueden* estar equivocados los resultados del experimento? Sólo siendo poco precisos. Por ejemplo, parece que la masa de un objeto nunca cambia: un trompo que gira tiene la misma masa que cuando está en reposo. Así se inventó una "ley": la masa es constante e independiente de la velocidad. Se sabe ahora que esa "ley" es incorrecta. Se encuentra que la masa aumenta con la velocidad, pero aumentos apreciables requieren velocidades cercanas a la de la luz. Una ley *verdadera* es: si un objeto se mueve con rapidez menor que 100 millas por segundo, la masa es constante dentro de una parte por millón. De esta manera aproximada ésta es una ley correcta. Se podría pensar que en la práctica esta nueva ley no da diferencia apreciable. Bueno, sí y no. Para velocidades ordinarias podemos olvidarla por cierto y usar la ley de masa constante como una buena aproximación. Pero para velocidades altas nos equivocamos y mientras mayor sea la velocidad tanto más equivocados estamos.

Por último, y esto es lo más interesante, *filosóficamente estamos completamente equivocados* con la ley aproximada. Nuestra imagen completa del mundo debe alterarse incluso si la masa cambia solamente un poco. Esto es un asunto muy peculiar de la filosofía o de las ideas que hay detrás de las leyes. Aun un efecto muy pequeño requiere a veces profundos cambios en nuestras ideas.

Ahora bien, ¿qué debemos enseñar primero? ¿Debemos enseñar la ley *correcta*, pero poco familiar con sus ideas extrañas y conceptualmente difíciles, por ejemplo, la teoría de la relatividad, el espacio-tiempo cuadrimensional, etcétera? ¿O debemos enseñar primero la ley sencilla de "masa constante", que es sólo aproximada, pero no incluye ideas tan difíciles? Lo primero es más excitante, más maravilloso y más entretenido, pero lo segundo es más fácil de comprender, y es el primer paso para un verdadero entendimiento de la primera idea. Este punto aparece repetidas veces al enseñar física. En diferentes ocasiones debemos resolverlo de maneras diferentes, pero en cada paso vale la pena aprender lo que se sabe ahora, hasta qué punto es exacto, cómo se ajusta a todo lo demás y cómo podrá cambiarse cuando aprendamos más.

Sigamos ahora con nuestro bosquejo, o mapa general, de lo que sabemos hoy día de ciencia (en particular de la física, pero también de otras ciencias que están en la periferia), de manera que cuando nos concentremos más tarde en algún punto particular tengamos alguna idea de los fundamentos, por qué ese punto particular resulta interesante y cómo se ajusta dentro de la gran estructura. De esta manera, ¿cuál *es* nuestra visión general del mundo?

1-2 La materia está formada de átomos

Si en algún cataclismo fuera destruido todo el conocimiento científico y solamente pasara una frase a la generación siguiente de criaturas, ¿cuál enunciado contendría el máximo de información en el mínimo de palabras? Yo creo que es la *hipótesis atómica* (o el *hecho* atómico, o como quieran llamarlo), que *todas las cosas están formadas por átomos, pequeñas partículas que se mueven con movimiento perpetuo, atrayéndose unas a otras cuando están separadas por una pequeña distancia, pero repeliéndose cuando se las trata de apretar una contra otra.* En esa sola frase, verán ustedes, hay una cantidad *enorme* de información referente al mundo, si se aplica sólo un poco de imaginación y pensamiento.

Para ilustrar el poder de la idea atómica, supongamos que tenemos una gota de agua de medio centímetro de lado. Si la miramos muy de cerca, no vemos más que agua —agua pareja y continua—. Incluso si la aumentamos con el mejor microscopio óptico disponible —aproximadamente 2 000 veces— entonces la gota de agua va a tener unos 12 metros de tamaño, casi tan grande como una pieza grande, y si miráramos de cerca, veríamos *todavía* agua relativamente pareja, pero de vez en cuando cuerpos parecidos a una pelota de futbol nadando de aquí

para allá. Muy interesante. Éstos son paramecios. Pueden detenerse en este punto y ponerse tan curiosos acerca de los paramecios con sus vibrantes cilios y cuerpos en contorsión, que no van a ir más allá, excepto tal vez para aumentar el tamaño del paramecio aún más y ver su interior. Esto, desde luego, es un tema de la biología, pero por el presente la dejamos y miramos con mayor atención aún el material del agua misma, aumentándolo nuevamente 2 000 veces. Ahora la gota de agua se extiende por unos 24 kilómetros y si la miramos muy de cerca vemos una especie de hormigueo, algo que ya no tiene una apariencia pareja; se parece a una multitud en un partido de futbol visto desde bastante distancia. Para ver qué es este hormigueo, lo aumentamos 250 veces más todavía y veremos algo parecido a lo que nos muestra la figura 1-1. Ésta es una representación del agua aumentada unas 1 000 millones de veces, pero idealizada en diferentes aspectos. En primer lugar, las partículas están dibujadas de manera sencilla con bordes definidos, lo cual no es exacto. Segundo, por simplicidad están bosquejadas en forma casi esquemática en un arreglo de dos dimensiones, pero se mueven desde luego en tres dimensiones. Nótese que hay dos tipos de "pompas" o círculos para representar los átomos de oxígeno (negro) e hidrógeno (blanco) y que cada oxígeno tiene dos hidrógenos unidos a él. (Cada pequeño grupo de un oxígeno con sus dos hidrógenos se denomina molécula). El dibujo está idealizado más aún en el sentido de que las verdaderas partículas en la naturaleza vibran y rebotan continuamente, rotando y contorsionándose la una alrededor de la otra. Deben imaginarse esto más bien como una representación dinámica y no estática. Otra cosa que no puede representarse en un dibujo es que las partículas están "pegadas entre sí" (que se atraen unas a otras, ésta tirada por aquélla, etcétera). El grupo completo está "encolado en un conjunto", por decirlo así. Por otra parte, las partículas no pueden atravesarse unas a otras. Si tratan de juntar dos de ellas demasiado cerca se repelen.

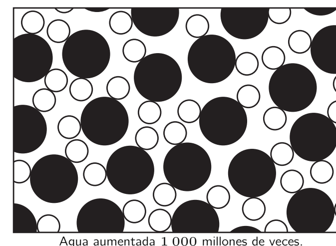

Agua aumentada 1 000 millones de veces.

Figura 1-1

Los átomos tienen un radio de 1 o 2×10^{-8} cm. Ahora bien, 10^{-8} cm se denomina un *ángstrom* (un nombre como cualquier otro), de manera que decimos que tienen un radio de uno o dos ángstroms (Å). Otra manera de recordar su tamaño es ésta: si una manzana se aumenta al tamaño de la Tierra, entonces los átomos de la manzana son aproximadamente del tamaño de la manzana original.

Ahora imagínense esta gran gota de agua, con todas estas partículas agitándose unidas entre sí y moviéndose lentamente en conjunto. El agua mantiene su volumen; no se deshace debido a la atracción mutua entre las moléculas. Si la gota está en una pendiente, donde se puede mover de un lugar a otro, el agua escurrirá, pero no desaparece simplemente —las cosas no se deshacen así tan fácilmente—, debido a la atracción molecular. Ahora bien, el movimiento de agitación es lo que nosotros representamos por *calor*: cuando aumentamos la temperatura, aumentamos el movimiento. Si calentamos el agua, la agitación aumenta y se incrementa el volumen entre los átomos, y si el calentamiento continúa, llega el momento en que la atracción entre las moléculas ya no es suficiente para mantenerlas unidas y ellas, *ahora sí*, vuelan en todas direcciones y se separan unas de otras. Desde luego, ésta es la manera como producimos vapor a partir del agua aumentando la temperatura —las partículas vuelan en todas direcciones debido al aumento del movimiento—.

En la figura 1-2 tenemos una representación del vapor. Esta representación del vapor falla en un aspecto: a presión atmosférica ordinaria ciertamente no habría más de tres moléculas de agua en esta figura. La mayoría de los cuadrados de este tamaño no contendrán ninguna, pero nosotros tenemos accidentalmente dos y media o tres en el dibujo (de esta manera no estará completamente vacío). Ahora bien, en el caso del vapor vemos las moléculas características con más claridad que en el caso del agua. Por simplicidad, las moléculas están dibujadas de manera que haya un ángulo de 120° entre los átomos de hidrógeno. En realidad el ángulo es de 105°3′, y la distancia entre el centro de un hidrógeno y el centro del oxígeno es de 0.957 Å, de manera que conocemos esta molécula bastante bien.

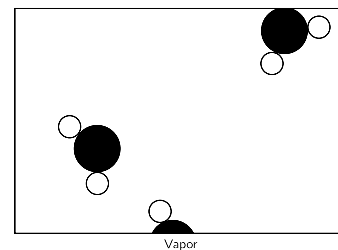

Vapor

Figura 1-2

Veamos cuáles son las propiedades del vapor de agua o de otro gas cualquiera. Las moléculas, estando separadas entre sí, van a rebotar contra las paredes.

Imagínense una pieza con un número de pelotas de tenis (unas 100 o algo así) rebotando en todas direcciones en movimiento perpetuo. Cuando bombardean la pared, esto empuja la pared apartándola. (Desde luego que tendremos que empujar la pared de vuelta). Esto quiere decir que el gas ejerce una fuerza a golpecitos que nuestros burdos sentidos (no habiendo sido aumentados nosotros mismos 1 000 millones de veces) sienten como un *empuje medio*. Para confinar un gas debemos aplicar una presión. La figura 1-3 muestra un recipiente común para contener gases (empleado en todos los textos), un cilindro con un pistón en él. Ahora bien, no importa cuál sea la forma de las moléculas de agua, de manera que por simplicidad las dibujaremos como pelotas de tenis o pequeños puntos. Estos objetos están en movimiento perpetuo en todas las direcciones. Así muchas de ellas están golpeando el pistón superior todo el tiempo, de modo que para evitar que sea botado fuera del tanque por este paciente y continuo golpeteo, debemos sujetar el pistón mediante una cierta fuerza que llamaremos *presión* (realmente, la presión multiplicada por el área es la fuerza). Está claro que la fuerza es proporcional al área, porque si incrementamos el área, pero mantenemos igual el número de moléculas por centímetro cúbico, aumentamos el número de colisiones con el pistón en la misma proporción que hemos incrementado el área.

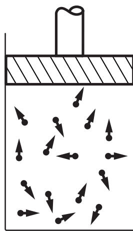


Figura 1-3

Coloquemos ahora doble cantidad de moléculas en el tanque, con el fin de duplicar la densidad, y que tengan la misma velocidad, esto es, la misma temperatura. Entonces, con bastante aproximación, el número de colisiones será doble, y como cada una de ellas es tan "energética" como antes, la presión es proporcional a la densidad. Si consideramos la verdadera naturaleza de las fuerzas entre los átomos, podemos esperar un pequeño decrecimiento de la presión debido a la atracción entre los átomos, y un pequeño aumento debido al volumen finito que ocupan. Sin embargo, con una excelente aproximación, si la densidad es suficientemente baja para que no haya demasiados átomos, *la presión es proporcional a la densidad*.

Podemos ver también algo más: si aumentamos la temperatura sin cambiar la densidad del gas, esto es, si aumentamos la velocidad de los átomos, ¿qué le sucederá a la presión? Bueno, los átomos pegan más fuerte porque se mueven más rápido, y además pegan más a menudo, de manera que la presión aumenta. Vean lo sencillas que son las ideas de la teoría atómica.

Consideremos ahora otra situación. Supongamos que el pistón se mueve hacia adentro, de modo que los átomos son comprimidos lentamente a un espacio más pequeño. ¿Qué sucede si un átomo choca contra el pistón en movimiento? Evidentemente toma velocidad de la colisión. Pueden comprobarlo haciendo rebotar una pelota de ping-pong contra una paleta que se mueve hacia adelante, por ejemplo, y hallarán que sale con más velocidad con la que chocó. (Ejemplo especial: si un átomo resultara estar en reposo y el pistón lo golpea, ciertamente se moverá). De esta manera los átomos resultan más "calientes" cuando se alejan del pistón que cuando lo chocaron. Por lo tanto, todos los átomos del recipiente habrán adquirido más velocidad. Esto quiere decir que *cuando comprimimos lentamente un gas, la temperatura del gas aumenta*. Así que, bajo una lenta *compresión*, un gas *aumentará* su temperatura, y bajo una lenta *expansión disminuirá* la temperatura.

Retornamos ahora a nuestra gota de agua y miramos en otra dirección. Supongamos que bajamos la temperatura de nuestra gota de agua. Suponga que la agitación de las moléculas de los átomos en el agua está continuamente decreciendo. Sabemos que existen fuerzas de atracción entre los átomos, de modo que después de un rato ya no estarán en condiciones de agitarse tan bien. Lo que sucederá a temperaturas muy bajas queda indicado en la figura 1-4: las moléculas se acomodan en una nueva estructura que es el *hielo*. Este diagrama esquemático particular del hielo está equivocado porque está dibujado en dos dimensiones, pero cualitativamente está bien. El punto interesante es que el material tiene un *lugar definido para cada átomo* y se puede apreciar fácilmente que si de una manera u otra podemos mantener todos los átomos de un extremo de la gota en una cierta ordenación, cada átomo en un cierto lugar, entonces debido a la estructura de interconexiones, que es rígida, el otro extremo a kilómetros de

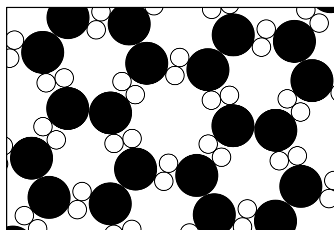


Figura 1-4

distancia (en nuestra escala aumentada) tendrá también una ubicación definida. Así, si sostenemos una aguja de hielo de un extremo, el otro extremo se resiste a ser desplazado, a diferencia del caso del agua, donde la estructura está destruida debido al aumento de la agitación, de manera que todos los átomos se mueven en forma diferente. La diferencia entre sólidos y líquidos es entonces que en un sólido los átomos están ordenados en un cierto tipo de estructura, llamada *estructura cristalina*, y no tienen una posición al azar a gran distancia; la posición de los átomos en un lado del cristal queda determinada por la de los otros átomos a distancia de millones de átomos al otro lado del cristal. La figura 1-4 es una ordenación inventada para el hielo, y a pesar de contener muchas de las características correctas del hielo, no es la verdadera ordenación. Uno de los aspectos correctos es que hay una parte de la simetría que es hexagonal. Pueden ver que si giran el dibujo en 60° alrededor de su eje, el dibujo resulta igual. Por lo tanto, hay una *simetría* en el hielo que es responsable del aspecto de seis caras de los copos de nieve. Otro asunto que podemos ver de la figura 1-4 es por qué el hielo se contrae cuando se derrite. La ordenación cristalina particular del hielo mostrada aquí tiene muchos "huecos" dentro de ella, como la estructura verdadera del hielo. Cuando se desmorona la estructura, estos huecos pueden ser ocupados por moléculas. La mayoría de las sustancias simples, con la excepción del agua y el metal de tipos de imprenta, se *dilatan* al fundirse, debido a que los átomos están muy compactos en el cristal sólido y al fundirse necesitan mayor espacio para agitarse, pero una estructura abierta sufre un colapso, como en el caso del agua.

Ahora bien, a pesar de que el hielo tiene una forma cristalina "rígida", su temperatura puede cambiar —el hielo tiene calor—. Si queremos, podemos cambiar su cantidad de calor. ¿Qué es el calor en el caso del hielo? Los átomos no están quietos; están agitándose y vibrando. De manera que, a pesar de existir un orden definido en el cristal —una estructura definida—, todos los átomos están vibrando "en posición". Si aumentamos la temperatura, vibran con amplitud cada vez mayor, hasta que finalmente se sacuden a sí mismos fuera de su posición. A esto lo llamamos *fusión*. Si decrecemos la temperatura, las vibraciones decrecen y decrecen hasta que, en el cero absoluto, los átomos quedan con una cantidad mínima de vibración, pero *no cero*. Este valor mínimo de movimiento que pueden tener los átomos no es suficiente para fundir las sustancias, con una excepción: el helio. El helio solamente decrece el movimiento atómico lo más que puede, pero aun en el cero absoluto todavía queda suficiente movimiento como para evitar su congelación. El helio, aun en el cero absoluto, no se congela, a menos que la presión se haga tan grande que los átomos se aplasten entre sí. Si aumentamos la presión, lo *podemos* hacer solidificar.

1-3 Procesos atómicos

Ya hablamos bastante de la descripción de los sólidos, líquidos y gases desde el punto de vista atómico. Sin embargo, la hipótesis atómica también describe *procesos*, y por eso veremos ahora algunos procesos desde el punto de vista atómico. El primer proceso que veremos está asociado con la superficie del agua. ¿Qué sucede en la superficie del agua? Haremos ahora el cuadro más complicado —y más realista— imaginándonos que la superficie está al aire. La figura 1-5 muestra la superficie de agua al aire. Vemos las moléculas de agua como antes, formando un cuerpo de agua líquida, pero ahora también vemos la superficie del agua. Sobre la superficie encontramos una serie de cosas: ante todo hay moléculas de agua, como en vapor. Éste es el *vapor de agua*, que siempre se encuentra sobre agua líquida. (Existe un equilibrio entre el vapor y el agua que será descrito posteriormente). Además encontramos otras moléculas —aquí dos átomos de oxígeno pegados entre sí—, formando una *molécula de oxígeno*, allá dos átomos de nitrógeno también pegados para formar una molécula de nitrógeno. El aire está formado casi enteramente por nitrógeno, oxígeno, algo de vapor de agua y cantidades menores de dióxido de carbono, argón y otras cosas. Así que sobre la superficie del agua está el aire, un gas, que contiene algo de vapor de agua.

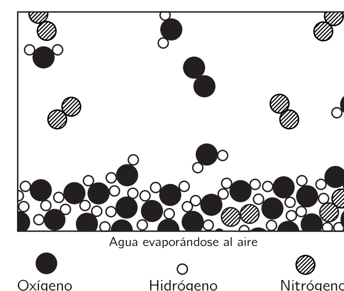
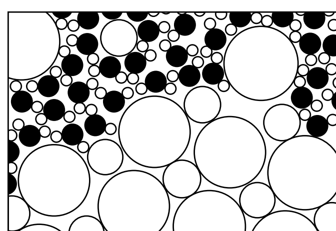


Figura 1-5

¿Qué sucede ahora en este cuadro? Las moléculas de agua están continuamente en agitación. De vez en cuando sucede que una de la superficie recibe un golpe un poco más fuerte que lo usual y sale. Es difícil ver este suceso en el dibujo, porque es un dibujo *estático*. Pero podemos imaginar que una molécula cerca de la superficie acaba de ser golpeada y está volando hacia afuera o que tal vez otra más ha sido afectada del mismo modo. De esta manera, molécula tras molécula, el agua desaparece —se evapora—. Pero si *cerramos* el recipiente por arriba, después de un tiempo hallaremos una gran cantidad de moléculas de agua entre las moléculas de aire. De vez en cuando, una de estas moléculas de vapor viene volando hacia abajo hacia el agua y queda aprisionada nuevamente. Vemos así que lo que parece ser una cosa muerta, sin interés —un vaso de agua con una tapa, que ha estado ahí quizá por 20 años—, realmente contiene un fenómeno dinámico e interesante que prosigue todo el tiempo. A nuestros ojos, nuestros imperfectos ojos, nada cambia, pero si pudiéramos verlo ampliado 1 000 millones de veces, veríamos que desde su propio punto de vista cambia continuamente: moléculas abandonan la superficie, moléculas regresan.

¿Por qué *no vemos cambio alguno*? ¡Porque tantas moléculas abandonan como regresan! A la larga “nada sucede”. Si ahora sacamos la tapa del recipiente y soplamos el aire húmedo a otra parte, remplazándolo por aire seco, el número de moléculas que abandonan el líquido sigue siendo el mismo como antes, porque esto depende de la agitación del agua, pero el número de las que regresan está fuertemente reducido porque hay tan pocas moléculas de agua sobre el agua. Por consiguiente, salen más de las que entran y el agua se evapora. ¡Luego, si desean evaporar agua, echen a andar el ventilador!

Aquí hay algo más. ¿Cuáles moléculas se van? Cuando una molécula se va es por una acumulación accidental extra de un poco más de energía de lo normal que ella necesita si debe escapar de la atracción de sus vecinas. Por lo tanto, ya que las que se van tienen más energía que el valor medio, las que se quedan tienen *menos* movimiento promedio de lo que tenían antes. De esta manera, el líquido gradualmente *se enfría* si se evapora. Desde luego, si una molécula de vapor baja desde el aire al agua, hay una gran atracción repentina a medida que la molécula se acerca a la superficie. Esto acelera la molécula que entra y da como resultado generación de calor. De modo, que cuando se van quitan calor; cuando regresan generan calor. Desde luego, cuando no hay evaporación neta, el resultado es nulo —el agua no cambia de temperatura—. Si soplamos el agua para mantener una preponderancia del número que se evapora, entonces el líquido se enfría. ¡Luego sople la sopa para enfriarla!



Sal disolviéndose en agua

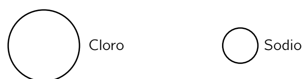
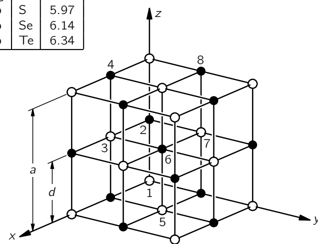


Figura 1-6

Desde luego deben darse cuenta que los procesos recién descritos son más complicados de lo que hemos indicado. No solamente pasa el agua al aire, sino que también de cuando en cuando una de las moléculas de oxígeno o nitrógeno entrará y “se perderá” en la masa de las moléculas de agua y se las arreglará para permanecer en el agua. De este modo el aire se disuelve en el agua; moléculas de oxígeno y nitrógeno se las arreglan para entrar al agua, y el agua contendrá aire. Si retiramos repentinamente el aire del recipiente, entonces las moléculas del gas saldrán más rápidamente de lo que entran y haciendo esto producirá burbujas. Esto es muy peligroso para los buzos, como ustedes sabrán.

Ahora pasaremos a otro proceso. En la figura 1-6 vemos, desde el punto de vista atómico, un sólido que se disuelve en agua. Si ponemos un cristal de sal en agua ¿qué sucederá? La sal es un sólido, un cristal, un sistema organizado de “átomos de sal”. La figura 1-7 es una ilustración de la estructura tridimensional de la sal común, cloruro de sodio. Estrictamente hablando, el cristal no está hecho de átomos, sino de lo que llamamos *iones*. Un ion es un átomo que tiene unos pocos electrones extra, o que ha perdido algunos electrones. En un cristal de sal encontramos iones cloro (átomos de cloro con un electrón extra) y iones sodio (átomos de sodio con un electrón de menos). Los iones están unidos unos a otros por atracción electrostática en la sal sólida, pero si los colocamos en agua, encontramos, debido a las atracciones del oxígeno negativo y del hidrógeno positivo por los iones, que algunos iones se sueltan. En la figura 1-6 vemos un ion cloro soltándose y otros átomos flotando en el agua en forma de iones. Este

Cristal	●	○	a (Å)
Sal común	Na	Cl	5.64
Silvina	K	Cl	6.28
	Ag	Cl	5.54
	Mg	O	4.20
Galena	Pb	S	5.97
	Pb	Se	6.14
	Pb	Te	6.34



Distancia al vecino más próximo $d = a/2$

Figura 1-7

dibujo fue hecho con algún cuidado. Nótese, por ejemplo, que los terminales de hidrógeno de las moléculas de agua están más próximos al ion cloro; en cambio, cerca del ion sodio encontramos más a menudo el terminal oxígeno, debido a que el sodio es positivo y el terminal oxígeno del agua es negativo, y se atraen eléctricamente. ¿Podemos decir con base en este dibujo si la sal se está *disolviendo en* agua o *cristalizando fuera* de ella? Desde luego que *no lo podemos* decir, porque mientras algunos de los átomos abandonan el cristal, otros átomos se vuelven a juntar con él. El proceso es *dinámico*, como en el caso de la evaporación, y depende de si hay más o menos sal en el agua que la cantidad necesaria para el equilibrio. Entendemos por equilibrio la situación en la cual la rapidez con que se van los átomos iguala exactamente a la rapidez con que regresan. Si casi no hubiera sal en el agua, se van más átomos de los que retornan, y la sal se disuelve. Si, por otro lado, hay demasiados "átomos de sal", regresan más de los que se van y la sal se cristaliza.

De paso, mencionaremos que el concepto de *molécula* de una sustancia es sólo aproximado y existe sólo para una cierta clase de sustancias. Está claro, en el caso del agua, que los tres átomos están realmente pegados unos a otros. No está tan claro en el caso del cloruro de sodio en el sólido. Hay una ordenación de iones sodio y cloro en el sólido. No hay manera natural de agruparlos como "moléculas de sal".

Retornando a nuestra discusión de solución y precipitación, si aumentamos la temperatura de la solución salina, aumenta la rapidez con que los átomos se retiran, y así también la rapidez con que los átomos retornan. Resulta, en general, muy difícil predecir en qué sentido se realiza, si se disuelve más o menos sólido. La mayoría de las sustancias se disuelven más, pero algunas sustancias se disuelven menos al aumentar la temperatura.

1-4 Reacciones químicas

En todos los procesos que han sido descritos hasta ahora, los átomos y los iones no han cambiado sus compañeros, pero desde luego hay circunstancias en las cuales los átomos cambian de combinación formando nuevas moléculas. Esto se ilustra en la figura 1-8. Un proceso en el cual ocurre un reordenamiento de los participantes atómicos es lo que llamamos *reacción química*. Los otros procesos descritos hasta ahora se llaman procesos físicos, pero no hay una distinción neta entre los dos. (La naturaleza no se preocupa de cómo los denominamos, simplemente sigue actuando). Ésta figura se supone que representa carbono que se quema en oxígeno. En el caso del oxígeno, *dos* átomos de oxígeno están unidos muy fuertemente. (¿Por qué no se unen *tres* o *cuatro*? Ésta es una de las características muy peculiares de tales procesos atómicos. Los átomos son muy especiales: les gustan ciertas estructuras particulares, ciertas direcciones particulares, etc. Es el trabajo de la física analizar por qué cada uno quiere lo que quiere. En todo caso, dos átomos de oxígeno forman, saturados y contentos, una molécula).

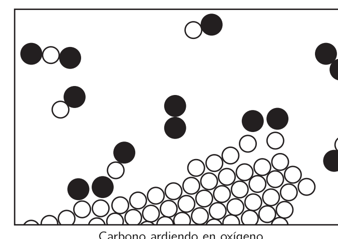

Carbono ardiendo en oxígeno

Figura 1-8

Se supone que los átomos de carbono están en un cristal sólido (que podría ser grafito o diamante).* Ahora, por ejemplo, una de las moléculas de oxígeno puede pasarse al carbono y cada átomo puede tomar un átomo de carbono y salir volando en una nueva combinación —"carbono-oxígeno"— que es una molécula del gas llamado monóxido de carbono. Se le da el nombre químico CO. Esto es muy sencillo: las letras "CO" son prácticamente el dibujo de aquella molécula. Pero el carbono atrae al oxígeno mucho más que el oxígeno al oxígeno o el carbono al carbono. Puede ser que el oxígeno llegue a este proceso con muy poca energía, pero el oxígeno y el carbono se unirán con tremenda violencia y conmoción y todo lo que está cerca de ellos captará esta energía. De este modo se genera una cantidad grande de energía de movimiento, energía cinética. Esto es, desde luego, *quemar*: obtener *calor* de la combinación del oxígeno y el carbono. El calor se tiene generalmente en forma de movimiento molecular del gas caliente,

* Se puede quemar un diamante en el aire.

pero en ciertas circunstancias puede ser tan enorme que genera *luz*. Así es como se obtienen las *llamas*.

Además, el monóxido de carbono no está satisfecho del todo. Le es posible ligar otro oxígeno, de manera que podemos tener una reacción mucho más complicada en la cual el oxígeno se combina con el carbono, mientras que al mismo tiempo se lleva a cabo una colisión con una molécula de monóxido de carbono. Podría unirse un átomo de oxígeno al CO y formar finalmente una molécula compuesta por un carbono y dos oxígenos, que se designa por CO_2 y que se llama dióxido de carbono. Si quemamos el carbono con muy poco oxígeno en una reacción muy rápida (por ejemplo, en un motor de automóvil, donde la explosión es tan rápida que no le queda tiempo para formar dióxido de carbono) se forma una cantidad considerable de monóxido de carbono. En muchos de estos nuevos arreglos se libera una cantidad bastante grande de energía, produciendo explosiones, llamas, etc., según la reacción. Los químicos han estudiado estas combinaciones de átomos, y han encontrado que todas las sustancias son algún tipo de *combinaciones de átomos*.

Para ilustrar esta idea, consideremos otro ejemplo. Si entramos en un campo de pequeñas violetas, sabemos cuál es su "olor". Es un cierto tipo de *molécula*, o combinación de átomos, que ha entrado en nuestras narices. Antes que nada, ¿*cómo* pudo llegar hasta allí? Es bastante sencillo. Si el olor es algún tipo de molécula en el aire, que se agita en todos los sentidos y siendo chocada a cada trecho, podría haber entrado *accidentalmente* en nuestras narices. Ciertamente no tiene ningún deseo particular de entrar en nuestra nariz. Es solamente una parte desvalida de un tumulto de moléculas, y en su vagar sin rumbo este pedazo de materia resulta que se encuentra en la nariz.

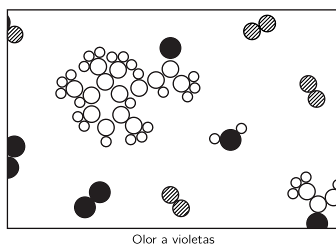


Figura 1-9

Ahora bien, los químicos pueden tomar moléculas particulares como el olor a violetas, analizarlas y decirnos cuál es el *ordenamiento exacto* de los átomos en el espacio. Sabemos que la molécula de dióxido de carbono es recta y simétrica: O—C—O. (Esto puede ser determinado fácilmente también por medios físicos). Sin embargo, aun para las combinaciones enormemente más complicadas de átomos que hay en la química, uno puede, mediante un largo y notable proceso de trabajo detectivesco, hallar la combinación de los átomos. La figura 1-9 representa el aire en la cercanía de una violeta; de nuevo encontramos nitrógeno y oxígeno en el aire, y vapor de agua. (¿Por qué hay vapor de agua? Porque la violeta está *húmeda*. Todas las plantas transpiran.) Pero también vemos un "monstruo" compuesto por átomos de carbono, hidrógeno y oxígeno, que han tomado una cierta forma especial para ordenarse. Es una ordenación mucho más complicada que la del dióxido de carbono; es, en efecto, una combinación enormemente complicada. Desgraciadamente no podemos representar todo lo que realmente se conoce de ella químicamente, porque la combinación precisa de todos los átomos se conoce en realidad en tres dimensiones, mientras que nuestro dibujo es solamente en dos. Los seis carbonos que forman un anillo no forman un anillo plano, sino una especie de anillo "arrugado". Todos sus ángulos y distancias son conocidos. De este modo una *fórmula* química es sólo una representación de la molécula. Cuando un químico escribe una de esas cosas en el pizarrón, trata de "dibujar", hablando llanamente, en dos dimensiones. Por ejemplo, nosotros vemos un "anillo" de seis carbonos y una "cadena" de carbonos colgando en el extremo, con un oxígeno en segundo lugar desde el extremo, tres hidrógenos unidos a un carbono, dos carbonos y tres hidrógenos asomados por aquí, etcétera.

¿Cómo descubre el químico cuál es la combinación? Mezcla botellas llenas de materiales y si se vuelve rojo, le dice que consiste de un hidrógeno y dos carbonos ligados aquí; si, por otra parte, se vuelve azul, se trata de un asunto totalmente diferente. Éste es uno de los trabajos detectivescos más fantásticos que se haya hecho nunca —la química orgánica—. Para descubrir el ordenamiento de los átomos en estas combinaciones tan enormemente complicadas, el químico se fija en lo que sucede si mezcla dos sustancias diferentes. El físico jamás creería que el químico sabe de lo que está hablando cuando describe las combinaciones de los átomos. Por unos 20 años ha sido posible en algunos casos observar estas moléculas (no tan complicadas como ésta, pero algunas que contienen parte

de ella) mediante un método físico, y ha sido posible localizar cada átomo, no mirando colores, sino *midiendo donde están.* ¡Y admírense!, los químicos casi siempre están en lo cierto.

Resulta, en efecto, que en el olor de las violetas hay tres moléculas levemente diferentes, que se diferencian solamente en la ordenación de los átomos de hidrógeno.

Un problema de la química es darle nombre a una sustancia, de manera que sepamos lo que es. ¡Encuentre un nombre para esta forma! El nombre no solamente debe dar idea de la forma, sino decir además que aquí hay un átomo de oxígeno, allá un hidrógeno (exactamente lo que es cada átomo y dónde está colocado). Así podemos apreciar que los nombres químicos deben ser complejos para que sean completos. Ustedes ven que el nombre de esta cosa en su forma más completa que les indique la estructura es 4-(2, 2, 3, 6 tetrametil-5-ciclohexenil)-3-buteno-2-ona, y eso les dice que ésta es su ordenación. Podemos apreciar las dificultades que tienen los químicos, y también apreciar las razones para usar nombres tan largos. ¡No es que ellos deseen ser oscuros, sino que tienen un problema extremadamente difícil al tratar de describir las moléculas en palabras!

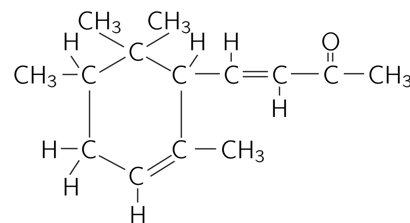


Figura 1-10. La sustancia representada es α-irona.

¿Cómo *sabemos* que existen los átomos? Mediante los trucos mencionados anteriormente: hacemos la *hipótesis* de que existen átomos, y un resultado tras otro sale como lo hemos predicho, como debería ser si las cosas *están* hechas de átomos. Hay además evidencias algo más directas, un buen ejemplo de lo cual es lo siguiente: los átomos son tan pequeños que no se pueden ver con un microscopio óptico —de hecho—, ni siquiera con un microscopio *electrónico.* (Con un microscopio óptico sólo pueden verse objetos que son mucho más grandes). Ahora, si los átomos están continuamente en movimiento, digamos en agua, y colocamos una pelota grande de algún material en el agua, una pelota mucho más grande que los átomos, la pelota se moverá en todas direcciones —muy parecido a un juego de pelota donde mucha gente trata de empujar una pelota muy grande—. La gente empuja en direcciones diferentes, y la pelota se mueve por el campo con una trayectoria irregular. Así, del mismo modo, la "pelota grande" se moverá por las desigualdades de las colisiones de un lado para el otro, de un instante al siguiente. Por eso, si miramos partículas muy pequeñas (coloides) en agua a través de un microscopio excelente, vemos una agitación perpetua de las partículas que es el resultado del bombardeo de los átomos. Esto se llama *movimiento browniano.*

Podemos ver otra evidencia más de la existencia de los átomos en la estructura de los cristales. En muchos casos las estructuras deducidas por análisis de rayos X están de acuerdo en sus "formas" espaciales con la forma que realmente presentan los cristales como se encuentran en la naturaleza. Los ángulos entre las diferentes "caras" de un cristal concuerdan, a menos de segundos de arco, con los ángulos deducidos suponiendo que un cristal está formado por muchas "capas" de átomos.

Todo está formado por átomos. Ésta es la hipótesis clave. La hipótesis más importante de toda la biología, por ejemplo, *es que todo lo que hacen los animales, lo hacen los átomos.* En otras palabras, *no hay nada que hagan los seres vivientes que no pueda ser entendido desde el punto de vista de que están hechos de átomos y éstos actúan según las leyes de la física.* Esto no se conocía desde un comienzo: fue necesaria alguna experimentación y teorización para sugerir esta hipótesis, pero ahora es aceptada, y es la teoría más útil para producir nuevas ideas en el campo de la biología.

Si un trozo de acero o un trozo de sal, que consisten de átomos uno junto al otro, pueden tener propiedades tan interesantes; si el agua —que no es sino estas pequeñas gotitas, kilómetros y kilómetros de lo mismo sobre la tierra— puede formar olas y espuma y hacer ruidos de torrente y figuras extrañas cuando corre sobre cemento; si todo esto, toda la vida de una corriente de agua, puede ser nada más que una pila de átomos, *¿cuánto más es posible?* Si en lugar de arreglar los átomos en alguna estructura definida, que se repita una y otra vez, siempre de nuevo, o incluso formando pequeños grupos complejos como en el olor de las violetas, hacemos una combinación que es *siempre diferente* de lugar a lugar, con distintos tipos de átomos combinados de muchas maneras

que cambian continuamente, no repitiéndose, ¿cuánto más maravillosamente es posible que este objeto se comporte? ¿Es posible que este "objeto" que camina de un lado a otro frente a ustedes hablándoles, sea un gran montón de estos átomos en un arreglo muy complejo, de modo que su mera complejidad hace vacilar la imaginación acerca de lo que puede hacer? Cuando decimos que somos un montón de átomos, no queremos decir que somos *meramente* un montón de átomos, porque un montón de átomos que no se repiten del uno al otro podría muy bien tener las posibilidades que ven frente a ustedes en el espejo.

2

Física básica

2-1 Introducción

En este capítulo examinaremos las ideas más fundamentales que tenemos sobre la física —la naturaleza de las cosas como las vemos en el presente—. No discutiremos la historia de cómo sabemos que todas estas ideas son verdaderas; ustedes aprenderán estos detalles a su debido tiempo.

Las cosas de las cuales nos preocupamos en ciencia se presentan en miríada de formas y con una multitud de atributos. Por ejemplo, si estamos parados en la playa y observamos el mar, vemos agua, el romper de las olas, la espuma, el movimiento chapoteante del agua, el sonido, el aire, los vientos y las nubes, también el Sol y el cielo azul y luz; hay allí arena y rocas de diferente dureza y duración, color y textura. Hay animales y algas, hambre y enfermedad, y el observador en la playa; puede haber aun felicidad y pensamientos. Cualquier otro lugar en la naturaleza tiene una variedad similar de cosas e influencias. Siempre será tan complicado como aquello, cualquiera que sea el lugar. La curiosidad exige que formulemos preguntas, que intentemos enlazar las cosas y tratemos de entender esta multitud de aspectos que resultan quizá de la acción de un número relativamente pequeño de cosas elementales y fuerzas que actúan en una variedad infinita de combinaciones.

Por ejemplo: ¿es la arena diferente a las rocas? Es decir, ¿es la arena quizá nada más que un gran número de piedras muy diminutas? ¿Es la Luna una gran roca? Si entendiéramos las rocas, ¿entenderíamos también la arena y la Luna? ¿Es el viento un chapoteo del aire análogo al movimiento chapoteante del agua en el mar? ¿Qué características comunes tienen los diferentes movimientos? ¿Qué es común a las diferentes clases de sonido? ¿Cuántos colores diferentes existen? Y así en tantas otras cosas. De esta manera tratamos de analizar gradualmente todas las cosas, de enlazar cosas que a primera vista parecen diferentes, con la esperanza de poder *reducir* el número de cosas *diferentes* y de esta manera comprenderlas mejor.

Hace algunos cientos de años se estableció un método para encontrar respuestas parciales a estas interrogantes. *Observación, razonamiento* y *experimentación* constituyen lo que llamamos el *método científico.* Tendremos que limitarnos a una mera descripción de nuestro punto de vista básico de lo que a veces se llama *física fundamental*, o ideas fundamentales que han surgido de la aplicación del método científico.

¿Qué queremos decir por "comprender" algo? Podemos imaginar que este conjunto complicado de cosas en movimiento, que constituyen "el mundo", es algo como un enorme juego de ajedrez jugado por dioses y nosotros somos observadores del juego. No conocemos las reglas del juego; lo único que nos está permitido es *observar* el juego. Por cierto, si observamos el tiempo suficiente, podemos eventualmente darnos cuenta de alguna de las reglas. Son *las reglas del juego* lo que entendemos por *física fundamental.* Sin embargo, aun si entendiéramos todas las reglas, podríamos no estar en condiciones de comprender por qué se hace un movimiento particular en el juego, meramente porque es demasiado complicado y nuestras mentes son limitadas. Si ustedes juegan ajedrez, deberán saber que es fácil aprender todas las reglas, y aun así es a veces muy difícil seleccionar la mejor jugada o comprender por qué un jugador mueve de una manera determinada. Así es en la naturaleza, sólo que en mayor grado; pero en última instancia podríamos encontrar todas las reglas. En realidad, no tenemos ahora todas las reglas. (Alguna vez en algún momento puede suceder algo como

el enroque que todavía no comprendemos.) Aparte de no conocer todas las reglas, lo que podemos explicar realmente a partir de aquellas reglas es muy limitado, porque casi todas las situaciones son tan enormemente complicadas que no podemos seguir las jugadas usando las reglas, mucho menos decir qué sucederá a continuación. Debemos, por lo tanto, limitarnos al problema más básico de las reglas del juego. Si conocemos las reglas, consideramos que "comprendemos" el mundo.

¿Cómo podemos decir si las reglas que "adivinamos" son realmente correctas, si no podemos analizar el juego muy bien? Existen, hablando a grandes rasgos, tres maneras. Primero, pueden existir situaciones donde la naturaleza se las ha arreglado, o nosotros arreglamos la naturaleza, para ser simple y tener tan pocas partes que podamos predecir exactamente lo que va a suceder, y de esta manera comprobar cómo funcionan nuestras reglas. (En una esquina del tablero puede haber solamente unas pocas piezas de ajedrez en juego y eso lo podemos descifrar exactamente.)

Una segunda manera útil para comprobar reglas es en términos de reglas menos específicas derivadas de aquéllas. Por ejemplo, la regla del movimiento de un alfil sobre un tablero de ajedrez es que se mueve solamente en diagonal. Se puede deducir, sin importar cuántos movimientos se hayan realizado, que un cierto alfil estará siempre sobre un cuadro blanco. Así, sin estar en condiciones de seguir los detalles, podemos comprobar siempre nuestra idea acerca del movimiento del alfil, averiguando si está siempre sobre un cuadro blanco. Por supuesto que estará allí por largo tiempo, hasta que de súbito encontramos que está sobre un cuadro *negro.* (Lo que sucedió, por supuesto, es que en el ínterin fue capturado, otro peón atravesó para ser coronado y se transformó en un alfil sobre un cuadro negro.) Así son las cosas en física. Por mucho tiempo tendremos una regla que funciona excelentemente en forma global, incluso cuando no podemos captar los detalles, y luego en algún momento podemos descubrir una *nueva regla.* Desde el punto de vista de la física básica, los fenómenos más interesantes se encuentran, por supuesto, en las partes *nuevas*, las partes donde las reglas no funcionan —¡no las partes donde *realmente* funcionan!—. Es de esta manera como descubrimos reglas nuevas.

La tercera manera de decir si nuestras ideas son correctas es relativamente tosca, pero probablemente la más poderosa de todas. Esto es, en forma *aproximada.* Aunque no seamos capaces de decir por qué Alekhine mueve *esa pieza particular*, quizá podamos entender *de algún modo* que está juntando sus piezas alrededor del rey para protegerlo, más o menos, ya que es lo sensato hacer en estas circunstancias. De la misma manera podemos a menudo comprender la naturaleza, más o menos, sin que seamos capaces de ver qué hace *cada pequeña pieza* en términos de nuestra comprensión del juego.

En un comienzo los fenómenos de la naturaleza se dividieron aproximadamente en clases, como el calor, la electricidad, la mecánica, el magnetismo, las propiedades de sustancias, los fenómenos químicos, la luz o la óptica, los rayos X, la física nuclear, la gravitación, los fenómenos mesónicos, etc. Sin embargo, el objetivo es ver la *naturaleza completa* como diferentes aspectos de un *conjunto único* de fenómenos. Éste es el problema básico de la física teórica de hoy —*encontrar las leyes que hay detrás del experimento; amalgamar estas clases*—. Históricamente hemos sido capaces siempre de amalgamarlas, pero con el transcurso del tiempo se encuentran cosas nuevas. Estábamos amalgamando muy bien, cuando de repente se descubrieron los rayos X. Entonces amalgamamos un poco más y se encontraron los mesones. Por lo tanto, en cada etapa del juego éste parece siempre bastante confuso. Mucho está amalgamado, pero siempre existen muchos alambres o hilos colgando en todas las direcciones. Ésta es hoy la situación que trataremos de describir.

Algunos ejemplos históricos de amalgamación son los siguientes. Tómese primero el *calor* y la *mecánica.* Cuando los átomos están en movimiento, mientras más movimiento más calor contiene el sistema y así *el calor y todos los efectos térmicos pueden representarse por las leyes de la mecánica.* Otra tremenda amalgamación fue el descubrimiento de la relación entre la electricidad, el magnetismo

y la luz, que se encontró que eran aspectos diferentes de una misma cosa, que llamamos hoy día el *campo electromagnético*. Otra amalgamación es la unificación de los fenómenos químicos, de las diversas propiedades de las diferentes sustancias y el comportamiento de las partículas atómicas, que está en la *mecánica cuántica de la química*.

La pregunta es, por supuesto, ¿será posible amalgamar *todo* y descubrir sólo que este mundo representa aspectos diferentes de *una* cosa? Nadie lo sabe. Todo lo que sabemos es que, si proseguimos, encontramos que podemos amalgamar partes y entonces encontramos algunas partes que no ajustan y seguimos tratando de armar el rompecabezas. Si existe un número finito de partes y aun si existe un límite para el rompecabezas, es, por supuesto, desconocido. Nunca se sabrá, hasta que hayamos terminado el cuadro, si es que lo logramos. Lo que deseamos hacer aquí es ver hasta qué punto ha avanzado este proceso de amalgamación, y cuál es la situación presente en la comprensión de fenómenos básicos en términos del conjunto más reducido de principios. Para expresarlo de una manera más simple, *¿de qué están hechas las cosas y cuántos elementos hay?*

2-2 La física antes de 1920

Es un poco difícil empezar de inmediato con el panorama presente; así, pues, veremos primero cómo se veían las cosas alrededor de 1920 y después extraeremos algunas cosas de esta imagen. Antes de 1920 nuestra imagen del mundo era algo como esto: el "escenario" en que el universo se desarrolla es el *espacio* tridimensional de la geometría, como fue descrito por Euclides, y las cosas cambian en un medio llamado *tiempo*. Los elementos en el escenario son *partículas*, por ejemplo, los átomos, que tienen algunas *propiedades*. Primero, la propiedad de la inercia: si una partícula se mueve, sigue moviéndose en la misma dirección, a menos que *fuerzas* actúen sobre ella. El segundo elemento lo constituyen entonces las *fuerzas*, que entonces se creía eran de dos variedades: primero un tipo enormemente complicado y detallado de fuerza de interacción que mantiene los diversos átomos en diferentes combinaciones de una manera complicada, que determinaba si la sal se disolvía más rápida o más lentamente cuando aumentamos la temperatura. La otra fuerza que se conocía era una interacción de largo alcance —una atracción suave y tranquila— que variaba inversamente con el cuadrado de la distancia y que se llamó *gravitación*. Esta ley se conocía y era muy simple. *Por qué* las cosas se mantienen en movimiento cuando se están moviendo o *por qué* existe una ley de gravitación, era desconocido, por supuesto.

Una descripción de la naturaleza es lo que nos interesa aquí. Desde este punto de vista, entonces, un gas y por cierto *toda* la materia, es una miríada de partículas en movimiento. De esta manera, muchas de las cosas que hemos visto mientras permanecíamos de pie a la orilla del mar pueden ser relacionadas inmediatamente. Primero la presión: ésta proviene de los choques de los átomos con las paredes, o lo que sea; el desplazamiento de los átomos, si en general se mueven todos en una dirección, es el viento; los movimientos internos al *azar* son el *calor*. Existen ondas de exceso de densidad, donde se han juntado demasiadas partículas, y así, a medida que se descomprimen, provocan más allá apilamientos de partículas y así sucesivamente. Esta onda de exceso de densidad es el *sonido*. Es un logro tremendo ser capaz de entender tanto. Algunas de estas cosas fueron descritas en el capítulo anterior.

¿Qué *clase* de partículas existen? En aquel tiempo consideraban que existían 92: a la larga se encontraron 92 clases diferentes de átomos. Tenían nombres diferentes asociados con sus propiedades químicas.

La parte siguiente del problema era, *¿qué son las fuerzas de corto alcance?* ¿Por qué el carbono atrae un oxígeno, o quizá dos oxígenos, pero no tres oxígenos? ¿Cuál es el mecanismo de interacción entre los átomos? ¿Será la gravitación? La respuesta es no. La gravedad es demasiado débil. Pero imaginemos una fuerza análoga a la gravedad, que varía inversamente con el cuadrado de la distancia, pero enormemente más poderosa y que tiene una diferencia. En la gravedad, todo

atrae a todo, pero ahora imaginemos que existen *dos clases* de "cosas", y que esta nueva fuerza (que es, por supuesto, la fuerza eléctrica) tiene la propiedad que los iguales *se repelen* y que los desiguales *se atraen*. La "cosa" que lleva esta interacción fuerte se llama *carga*.

Entonces, ¿qué es lo que tenemos? Supóngase que tenemos dos desiguales que se atraen entre sí, uno positivo y uno negativo, y que quedan unidos a una distancia muy corta. Supóngase que tenemos otra carga a cierta distancia. ¿Sentirá ella alguna atracción? *Prácticamente* no sentirá *ninguna*, porque si las dos primeras son iguales en magnitud, la atracción de una y la repulsión de la otra se compensan. Por lo tanto, hay muy poca fuerza a cualquier distancia apreciable. Por otra parte, si nos *acercamos mucho* con la carga extra, surge *atracción*, porque la repulsión de las iguales y la atracción de las desiguales tenderán a acercar las desiguales y a separar las iguales. Entonces la repulsión será *menor* que la atracción. Ésta es la razón de que los átomos que están constituidos por cargas eléctricas positivas y negativas sienten una fuerza muy débil cuando están separados por una distancia apreciable (aparte de la gravedad). A medida que se acercan, cada una puede "mirar el interior" de la otra y reordenar sus cargas, con el resultado de que tienen una interacción muy fuerte. La base última de la interacción entre los átomos es *eléctrica*. Ya que esta fuerza es tan enorme, todos los positivos y todos los negativos se juntarán normalmente en una combinación tan íntima como sea posible. Todas las cosas, aun nosotros mismos, están hechas de partes positivas y negativas finamente granuladas que interactúan de manera muy fuerte, todas perfectamente compensadas. A veces, por accidente, podemos quitar por frotamiento unos pocos negativos o unos pocos positivos (corrientemente es más fácil quitar los negativos) y en esas circunstancias encontramos la fuerza eléctrica *no compensada* y entonces podemos ver los efectos de estas atracciones eléctricas.

Para dar una idea de cuánto más fuerte es la electricidad que la gravitación, considérese dos granos de arena, de un milímetro de diámetro, separados 30 metros. Si la fuerza entre ellos no estuviera compensada, si todo atrae a todo, en vez de repelerse los iguales, de modo que no haya cancelación, ¿cuánta fuerza habrá? ¡Habrá una fuerza de *tres millones de toneladas* entre ambos! Vean que hay muy, pero *muy* poco exceso o déficit en el número de cargas negativas o positivas necesarias para producir efectos eléctricos apreciables. Ésta es, ciertamente, la razón de por qué ustedes no pueden ver la diferencia entre una cosa cargada eléctricamente y una sin carga, tan pocas partículas entran en juego que es difícil que produzcan una diferencia en el peso o en el tamaño de un objeto.

Con esta imagen los átomos eran más fáciles de comprender. Se imaginaba que tienen un "núcleo" en el centro, que está cargado electropositivamente y es muy pesado, y el núcleo está rodeado por cierto número de "electrones", que son sumamente livianos y están cargados negativamente. Ahora avanzamos un poco más en nuestra historia para hacer notar que en el núcleo mismo se encontraron dos clases de partículas, protones y neutrones, casi del mismo peso y muy pesados. Los protones están cargados eléctricamente y los neutrones son neutros. Si tenemos un átomo con seis protones en el interior de su núcleo y éste está rodeado por seis electrones (las partículas negativas en el mundo ordinario de la materia son todos electrones, y éstos son muy livianos comparados con los protones y neutrones que constituyen los núcleos), éste sería el número atómico seis en la tabla de la química, y se llama carbono. El número atómico ocho se llama oxígeno, etc., porque las propiedades químicas dependen de los electrones *externos* y, en realidad, sólo de *cuántos* electrones hay. Así las propiedades *químicas* de una sustancia dependen sólo de un número, el número de electrones. (La lista completa de elementos de los químicos podría haberse realmente denominado 1, 2, 3, 4, 5, etc. En vez de decir "carbono" podríamos decir "elemento seis", significando esto seis electrones; pero, por supuesto, cuando los elementos fueron descubiertos, no se sabía que pudieran ser numerados de esta manera, y en segundo lugar haría aparecer todo más bien complicado. Es mejor tener nombres y símbolos para estas cosas, más bien que indicar todo por un número.)

Se descubrió aún más acerca de las fuerzas eléctricas. La interpretación natural

de la interacción eléctrica es que dos objetos simplemente se atraen: positivo contra negativo. Sin embargo, se descubrió que esto era una idea inadecuada para representarla. La representación más adecuada de la situación es afirmar que la existencia de la carga positiva en cierto sentido distorsiona o crea una "condición" en el espacio tal que si introducimos una carga negativa, ella siente una fuerza. Esta potencialidad para producir una fuerza se llama *campo eléctrico*. Cuando colocamos un electrón en un campo eléctrico decimos que "tira de él". Tenemos entonces dos reglas: (a) las cargas producen un campo y (b) las cargas en campos experimentan fuerzas y se mueven. La razón para esto se aclarará cuando discutamos el siguiente fenómeno: si cargamos eléctricamente un cuerpo, digamos un peine, y después colocamos un trozo de papel cargado a cierta distancia y movemos el peine de atrás para adelante, el papel responderá apuntando siempre hacia el peine. Si lo agitamos más rápidamente se descubrirá que el papel se queda un poco más atrás, *hay un retardo* en la acción. (En la primera etapa, cuando movemos el peine más bien lentamente, encontramos una complicación que es el *magnetismo*. Las influencias magnéticas tienen que ver con *cargas en movimiento relativo*, por lo que las fuerzas magnéticas y eléctricas pueden ser realmente atribuidas a un campo, como dos aspectos diferentes de exactamente una misma cosa. Un campo eléctrico variable no puede existir sin magnetismo.) Si retiramos más el papel cargado, el retraso es mayor. Entonces se observa algo interesante. A pesar de que las fuerzas entre dos objetos cargados deberían variar inversamente con el *cuadrado* de la distancia, se encuentra que cuando agitamos la carga, la influencia se extiende *mucho más allá* de lo que podríamos suponer a primera vista. Esto es, el efecto decrece mucho más lentamente que la inversa del cuadrado.

Hay aquí una analogía: si estamos en una piscina y hay muy cerca un corcho flotando, podremos moverlo "directamente" empujando el agua con otro corcho. Si ustedes miran solamente ambos *corchos*, todo lo que verán es que uno se mueve inmediatamente en respuesta al movimiento del otro —existe algún tipo de *"interacción"* entre ellos—. Por cierto, lo que realmente hacemos es perturbar el *agua*; el *agua* perturba entonces el otro corcho. Podríamos establecer entonces una "ley": si empuja un poco el agua, un objeto próximo en el líquido se moverá. Si estuviera más alejado, por supuesto, el segundo corcho apenas se movería, porque movemos *localmente* el agua. Por otra parte, si agitamos el corcho, un nuevo fenómeno estará implicado, en el cual el movimiento del agua mueve el agua allí, etc., y se propagan *ondas*; así que por agitación hay una influencia de *mucho mayor alcance*; una influencia oscilatoria, que no puede ser comprendida a partir de la interacción directa. Por lo tanto, la idea de interacción directa debe remplazarse por la existencia del agua, o en el caso eléctrico, por lo que llamamos el *campo electromagnético*.

El campo electromagnético puede transportar ondas; algunas de estas ondas son *luz*; otras se usan en *radiodifusión*, pero el nombre general es *ondas electro-*

Tabla 2-1

El espectro electromagnético

Frecuencia en oscilaciones/s	Nombre	Comportamiento aproximado
10^2	Perturbación eléctrica	Campo
$5 \times 10^5 - 10^6$	Radiodifusión	Ondas
10^8	FM—TV	
10^{10}	Radar	
$5 \times 10^{14} - 10^{15}$	Luz	
10^{18}	Rayos X	Partículas
10^{21}	Rayos γ nucleares	
10^{24}	Rayos γ "artificiales"	
10^{27}	Rayos γ en rayos cósmicos	

magnéticas. Estas ondas oscilatorias pueden tener varias *frecuencias.* Lo único que es realmente diferente de una a otra onda es la *frecuencia des oscilación.* Si sacudimos una carga de aquí para allá más y más rápidamente y observamos los efectos, obtenemos toda una serie de diferentes tipos de efectos que están unificados por la especificación de un solo número, el número de oscilaciones por segundo. La "captación" común que obtenemos de las corrientes eléctricas de los circuitos en las paredes de un edificio tiene una frecuencia de alrededor de 100 ciclos por segundo. Si aumentamos la frecuencia a 500 o 1 000 kilociclos (un kilociclo = 1 000 ciclos) por segundo, estamos "en el aire", ya que éste es el intervalo de frecuencias que se usa para radiodifusión. (¡Por supuesto, ello no tiene nada que ver con el *aire*! Podemos tener radiodifusión sin aire alguno.) Si aumentamos nuevamente la frecuencia, llegamos al intervalo que es usado para FM y TV. Yendo todavía más allá, usamos ciertas ondas cortas, por ejemplo para *radar.* Aún más altas, ya no necesitamos un instrumento para "ver" la cosa, podemos verla con el ojo humano. En el intervalo de frecuencias de 5×10^{14} a 10^{15} ciclos por segundo nuestros ojos verían la oscilación del peine cargado, si lo pudiéramos sacudir así de rápidamente, como luz roja, azul o violeta, dependiendo de la frecuencia. Las frecuencias por debajo de este intervalo se llaman infrarroja y por encima, ultravioleta. El hecho de que podamos ver dentro de un intervalo particular de frecuencias no hace más impresionante esa parte del espectro electromagnético que las otras desde el punto de vista físico, pero desde el punto de vista humano, por supuesto, *es* más interesante. Si subimos aún más la frecuencia, obtenemos rayos X. Los rayos X no son otra cosa que luz de muy alta frecuencia. Si subimos todavía más, obtenemos rayos gamma. Estos dos términos, rayos X y rayos gamma, se usan casi en forma sinónima. Comúnmente los rayos electromagnéticos provenientes de los núcleos se llaman rayos gamma, mientras que los de alta energía de los átomos se llaman rayos, pero a la misma frecuencia son físicamente indistinguibles, cualquiera que sea su origen. Si vamos a frecuencias todavía más altas, digamos 10^{24} ciclos por segundo, encontramos que podemos producir esas ondas artificialmente, por ejemplo con el sincrotrón aquí en el Caltech. Podemos encontrar ondas electromagnéticas con frecuencias estupendamente altas —con oscilaciones aún 1 000 veces más rápidas— en las ondas encontradas en los *rayos cósmicos.* Estas ondas no pueden ser controladas por nosotros.

2-3 Física cuántica

Una vez descrita la idea del campo electromagnético y que este campo puede transportar ondas, pronto nos damos cuenta que estas ondas se comportan realmente de una manera extraña, que tienen una apariencia muy poco ondulatoria. ¡A frecuencias más elevadas se comportan mucho más como *partículas*! Es la *mecánica cuántica*, descubierta poco después de 1920, que explica este comportamiento extraño. En los años anteriores a 1920 la imagen del espacio tridimensional, y del tiempo como una cosa separada, fue cambiada por Einstein, primero en una combinación que llamamos espacio-tiempo, y después aún más en un espacio-tiempo *curvo*, para representar la gravitación. Así, el "escenario" se cambia a espacio-tiempo y la gravitación es presumiblemente una modificación del espacio-tiempo. Entonces se encontró también que las reglas para los movimientos de partículas eran incorrectas. Las reglas mecánicas de "inercia" y "fuerzas" están *equivocadas* —las leyes de Newton están *equivocadas*— en el mundo de los átomos. En cambio, se descubrió que las cosas a escala pequeña no se comportan *en absoluto* como las cosas a escala grande. Esto es lo que hace difícil la física —y muy interesante—. Es difícil porque la manera como se comportan las cosas a escala pequeña es tan "innatural"; no tenemos experiencia directa de eso. Aquí las cosas se comportan de un modo distinto a todo lo que conocemos; así que es imposible describir este comportamiento de ninguna otra manera que no sea la analítica. Esto es difícil y requiere mucha imaginación.

La mecánica cuántica tiene muchos aspectos. En primer lugar, la idea de que una partícula tiene una ubicación definida y una velocidad definida ya no

se acepta más; esto es erróneo. Para dar un ejemplo de lo errada que está la física clásica, existe una regla en la mecánica cuántica que afirma que no se puede saber simultáneamente dónde está algún objeto y a qué velocidad se mueve. La indeterminación del momentum y la indeterminación de la posición son complementarias y el producto de ambas está acotado por una constante pequeña. Podemos escribir la ley así: $\Delta x\,\Delta p \geq \hbar/2$, pero la explicaremos con más detalles más adelante. Esta regla es la explicación de una paradoja muy misteriosa: si los átomos están hechos de cargas positivas y negativas, ¿por qué las cargas negativas no se ubican simplemente sobre las cargas positivas (se atraen entre sí) y se acercan tanto como para anularse completamente? *¿Por qué son tan grandes los átomos?* ¿Por qué está el núcleo en el centro con los electrones a su alrededor? Se pensó en un comienzo que esto era porque el núcleo era tan grande; pero no, el núcleo es *muy pequeño*. Un átomo tiene un diámetro de alrededor de 10^{-8} cm. El núcleo tiene un diámetro de alrededor de 10^{-13} cm. Si tuviéramos un átomo y deseáramos ver el núcleo, tendríamos que aumentarlo hasta que todo el átomo fuera del tamaño de una pieza grande y entonces el núcleo no sería más que un granito, que apenas se podría distinguir con el ojo, pero casi *todo el peso* del átomo está en ese *núcleo* infinitesimal. ¿Qué impide a los electrones caer simplemente sobre él? Este principio: si estuvieran en el núcleo, conoceríamos exactamente su posición y el principio de incertidumbre requeriría entonces que ellos tuvieran un momentum muy *grande* (pero incierto), es decir, una *energía cinética* muy grande. Con esta energía se escaparían del núcleo. Así llegan a un acuerdo: se dejan un pequeño espacio para esta incertidumbre y entonces se agitan con un cierto movimiento mínimo, de acuerdo con esta regla. (Recuérdese que cuando se enfría un cristal al cero absoluto, decíamos que los átomos todavía se agitan, no se detienen. ¿Por qué? Si se detuvieran, sabríamos dónde se encuentran y que tienen movimiento nulo, lo cual está contra el principio de incertidumbre. No podemos saber dónde están y cuán rápido se mueven, ¡por lo tanto deben estar meneándose continuamente allí!)

Otro cambio del mayor interés en las ideas y la filosofía de las ciencias, introducido por la mecánica cuántica, es éste: no es posible predecir *exactamente* qué sucederá en cualquier circunstancia. Por ejemplo, es posible disponer un átomo de manera que esté a punto de emitir luz y podemos medir cuándo ha emitido luz, recogiendo una partícula fotónica, que describiremos pronto. Sin embargo, no podemos predecir *cuándo* emitirá la luz o, con varios átomos, *cuál es* el que lo hará. Ustedes podrían decir, que esto es porque hay algunos "engranajes" internos que no hemos mirado suficientemente de cerca. No, no *existen* engranajes internos; la naturaleza como la entendemos hoy se comporta de una manera tal, que es *fundamentalmente imposible* hacer una predicción precisa de *lo que va a suceder exactamente* en un experimento dado. Esto es una cosa horrible; en efecto, los filósofos habían dicho antes que uno de los requisitos fundamentales de la ciencia es que siempre que se ponen las mismas condiciones debe suceder lo mismo. Esto simplemente *no es cierto*, *no* es una condición fundamental de la ciencia. El hecho es que no sucede la misma cosa, que podemos encontrar sólo un promedio, estadísticamente, de lo que sucede. A pesar de todo, la ciencia no se ha derrumbado completamente. A propósito, los filósofos dicen mucho acerca de lo que es *absolutamente necesario* para la ciencia, y ello resulta siempre, hasta donde uno puede ver, más bien ingenuo y probablemente equivocado. Por ejemplo, alguno que otro filósofo decía que es fundamental para el logro científico que si un experimento se realiza, digamos en Estocolmo, y luego el mismo experimento se realiza en Quito, deben encontrarse los *mismos resultados*. Esto es totalmente falso. No es necesario que la *ciencia* haga eso; puede ser una *realidad de la experiencia*, pero no es necesario. Por ejemplo, si uno de los experimentos consiste en observar el cielo y ver las auroras boreales en Estocolmo, usted no las verá en Quito; aquél es un fenómeno diferente. "Pero —dirán ustedes—, esto es algo que tiene que ver con el exterior; ¿puede usted encerrarse en una caja en Estocolmo, bajar las cortinas y obtener alguna diferencia?" Por supuesto. Si tomamos un péndulo con una suspensión universal, lo desplazamos y lo soltamos, entonces el péndulo oscilará casi en un plano, pero no totalmente. De

forma lenta el plano irá cambiando en Estocolmo, pero no en Quito. Las celosías también están bajas. El hecho que esto ocurra no trae consigo la destrucción de la ciencia. ¿Cuál *es* la hipótesis fundamental de la ciencia, la filosofía fundamental? La establecimos en el primer capítulo: *la única prueba de la validez de cualquier idea es el experimento.* Si sucede que la mayoría de los experimentos se verifican lo mismo en Quito que en Estocolmo, entonces esta "mayoría de experimentos" se usará para formular alguna ley general, y de aquellos experimentos que no tengan el mismo resultado diremos que fueron el resultado del medio en Estocolmo. Inventaremos alguna manera de resumir los resultados del experimento y no necesitamos que se nos diga con anticipación qué aspecto tendrá esta manera. Si se nos dice que el mismo experimento va a dar siempre el mismo resultado, eso está todo muy bien; pero, si cuando lo ensayamos, *no* lo da, entonces *no* lo da. Debemos simplemente aceptar lo que vemos y, entonces, formular todo el resto de nuestras ideas en términos de nuestra experiencia real.

Volviendo nuevamente a la mecánica cuántica y la física fundamental, no podemos entrar en detalles de los principios cuánticos en este momento, por supuesto, porque éstos son bastante difíciles de comprender. Supondremos que existen y continuaremos describiendo cuáles son algunas de las consecuencias. Una de las consecuencias es que cosas que considerábamos como ondas, también se comportan como partículas y las partículas se comportan como ondas; en efecto, todo se comporta de la misma manera. No existe distinción entre ondas y partículas. Por lo tanto, la mecánica cuántica *unifica* la idea de campo y sus ondas y la de partículas, todo en una. Ahora bien, es cierto que si la frecuencia es baja, el aspecto de campo del fenómeno es más evidente, o más útil como una descripción más aproximada en términos de la experiencia diaria. Pero a medida que la frecuencia aumenta, los aspectos corpusculares del fenómeno se hacen más evidentes con el equipo con el que hacemos corrientemente las mediciones. En efecto, a pesar de que hemos mencionado muchas frecuencias, no se ha detectado aún ningún fenómeno en el que intervenga directamente una frecuencia por sobre aproximadamente 10^{12} ciclos por segundo. Solamente *deducimos* las frecuencias más altas a partir de la energía de las partículas, por medio de una regla que supone que la idea corpúsculo-onda de la mecánica cuántica es válida.

Así tenemos un nuevo punto de vista de la interacción electromagnética. Tenemos que agregar un nuevo tipo de *partícula* al electrón, al protón y al neutrón. La nueva partícula se llama *fotón.* El nuevo punto de vista de la interacción de electrones y fotones, que es la teoría electromagnética, pero con todo cuánticamente correcto, se llama *electrodinámica cuántica.* Esta teoría fundamental de la interacción de luz y materia, o campo eléctrico y cargas, es nuestro mayor éxito hasta ahora en física. En esta sola teoría tenemos todas las reglas básicas para todos los fenómenos ordinarios, excepto para la gravitación y los procesos nucleares. Por ejemplo, de la electrodinámica cuántica salen todas las leyes eléctricas, mecánicas y químicas conocidas: las leyes para la colisión de bolas de billar, el movimiento de alambres en campos magnéticos, el calor específico del monóxido de carbono, el color de los letreros de neón, la densidad de la sal y las reacciones de hidrógeno y oxígeno para producir agua, son todas consecuencias de esta sola ley. Todos estos detalles se pueden elaborar si la situación es lo suficientemente simple para que nosotros hagamos una aproximación, lo que no sucede casi nunca, pero a menudo podemos comprender más o menos lo que está sucediendo. En el presente, no se han encontrado excepciones a las leyes de la electrodinámica cuántica fuera del núcleo y no sabemos si hay una excepción ahí, porque simplemente no sabemos qué sucede en el núcleo.

En principio, entonces, la electrodinámica cuántica es la teoría de toda la química y de la vida, si la vida se reduce en última instancia a la química y por lo tanto precisamente a la física, porque la química está ya reducida (siendo ya conocida la parte de la física que está comprendida en la química). Más aún, la misma electrodinámica cuántica, esta cosa grandiosa, predice muchas cosas nuevas. En primer lugar da las propiedades de fotones de muy alta energía, rayos gamma, etc. Predijo otra cosa muy notable: además del electrón, debería haber otra partícula de la misma masa, pero de carga opuesta, llamada *positrón*, y

esas dos, al encontrarse, pueden aniquilarse entre sí, con la emisión de luz o rayos gamma. (Después de todo, luz y rayos gamma son lo mismo, son sólo puntos diferentes sobre una escala de frecuencias.) La generalización de esto, que para cada partícula existe una antipartícula, resulta ser cierta. En el caso de los electrones, la antipartícula tiene otro nombre —se llama positrón, pero para la mayoría de las otras partículas se llama anti-tal-o-cual, como antiprotón o antineutrón—. En electrodinámica cuántica se introducen *dos números* y se supone que de allí sale la mayoría de los otros números en el universo. Los dos números que se introducen se llaman la masa del electrón y la carga del electrón. En realidad, esto no es totalmente cierto, porque tenemos un conjunto completo de números para la química, que indican cuál es el peso de los núcleos. Esto nos conduce a la parte siguiente.

2-4 Núcleos y partículas

¿De qué están hechos los núcleos y cómo se mantienen? Se encuentra que los núcleos se mantienen unidos debido a fuerzas enormes. Cuando éstas se liberan, la energía liberada es tremenda comparada con la energía química, en la misma relación que la explosión de la bomba atómica con una explosión de TNT porque, por supuesto, la bomba atómica tiene que ver con cambios en el interior del núcleo, mientras que la explosión de TNT tiene que ver con cambios de los electrones en el exterior de los átomos. La pregunta es: ¿cuáles son las fuerzas que mantienen unidos los protones y neutrones en el núcleo? Tal como la interacción eléctrica puede ser relacionada con una partícula, un fotón, Yukawa sugirió que las fuerzas entre neutrones y protones también poseen un campo de alguna clase, y que cuando este campo se agita, se comporta como una partícula. Así podría haber algunas otras partículas en el universo además de protones y neutrones y él pudo deducir las propiedades de estas partículas a partir de las características ya conocidas de las fuerzas nucleares. Por ejemplo, él predijo que deberían tener una masa 200 o 300 veces la de un electrón; ¡y he ahí, en los *rayos cósmicos* se descubrió una partícula de la masa correcta! Pero más tarde resultó ser la partícula equivocada. Se la llamó muon.

Sin embargo, un poco más tarde, en 1947 o 1948 se encontró otra partícula, el mesón π, o pion, que satisfacía el criterio de Yukawa. Entonces, además del protón y del neutrón, debemos agregar el pion para obtener las fuerzas nucleares. Ahora ustedes dirán: "¡Qué grandioso!, con esta teoría hacemos la dinámica cuántica nuclear, usando los piones como quería hacerlo Yukawa, vemos si funciona y todo estará explicado". Mala suerte. Resulta que los cálculos implicados en esta teoría son tan difíciles, que nadie ha podido descifrar cuáles son las consecuencias de la teoría o comprobarla con el experimento, ¡y esto ha seguido así hasta ahora, durante casi 20 años!

Estamos así atascados con una teoría y no sabemos si está correcta o incorrecta, pero sí sabemos que está un *poco* equivocada, o por lo menos incompleta. Mientras nosotros hemos estado divagando teóricamente, tratando de calcular las consecuencias de esta teoría, los experimentadores han encontrado algunas cosas. Por ejemplo, ya habían encontrado este muon y todavía no sabemos dónde encaja. También en los rayos cósmicos se había encontrado un gran número de otras partículas "extra". Resulta que hoy día tenemos aproximadamente 30 partículas y es muy difícil comprender las relaciones entre todas estas partículas y para qué la naturaleza las quiere, o cuáles son las conexiones entre una y otra. No entendemos hoy estas varias partículas como aspectos diferentes de la misma cosa, y el hecho de que tengamos tantas partículas desconectadas es una representación del resultado de tener tanta información inconexa sin una buena teoría. Tras los grandes éxitos de la electrodinámica cuántica existe un cierto conocimiento de la física nuclear, que es un conocimiento crudo, una especie de mitad experiencia y mitad teoría, al suponer un tipo de fuerza entre protones y neutrones y viendo qué sucede, pero sin comprender realmente de dónde provienen las fuerzas. Aparte de esto, hemos hecho muy poco progreso. Hemos juntado un número enorme de elementos químicos. En el caso de la química, ha aparecido súbitamente

una relación inesperada entre estos elementos que está incorporada en la tabla periódica de Mendeléyev. Por ejemplo, el sodio y el potasio son casi iguales en sus propiedades químicas y se encuentran en la misma columna de la tabla de Mendeléyev. Hemos estado buscando una tabla tipo Mendeléyev para las nuevas partículas. Una tabla así para las partículas nuevas fue confeccionada independientemente por Gell-Mann en los Estados Unidos y Nishijima en el Japón. La base de su clasificación es un número nuevo, como la carga eléctrica, que puede ser asignado a cada partícula, llamado su "extrañeza", S. Este número se conserva, como la carga eléctrica, en reacciones que tienen lugar debido a fuerzas nucleares.

Tabla 2-2. Partículas elementales

Masa en MeV	Carga $-e$	Carga 0	Carga $+e$	Agrupamiento y extrañeza	
1 400	$Y_1^- \to \Lambda^0 + \pi^-$	$Y_1^0 \to \Lambda^0 + \pi^0$	$Y_1^+ \to \Lambda^0 + \pi^+$ 1 395	S=−2	Bariones
1 300	Ξ^- 1 319	Ξ^0 1 311		S=−2	Bariones
1 200	Σ^- 1 196	Σ^0 1 191	Σ^+ 1 189	S=−1	Bariones
1 100		Λ^0 1 115		S=−1	Bariones
1 000					
		n 939	p 938	S=0	Bariones
900					
800		$\omega^0 \to \pi + \pi + \pi$		S=0	Mesones
	$\rho^- \to \pi + \pi$	$\rho^0 \to \pi + \pi$	$\rho^\pm \to \pi + \pi$	S=0	Mesones
700					
600					
500	K^- 494	$K^0 \overline{K}^0$ 498	K^+ 494	S=±1	Mesones
400					
300					
200					
	π^- 139.6	π^0 135.0	π^+ 139.6	S=0	Mesones
100	μ^- 105.6				Leptones
0	e^- 0.51	ν^0 0			Leptones

En la tabla 2-2 se han indicado todas las partículas. No podemos discutirlas mucho en esta etapa, pero por lo menos la tabla les muestra lo mucho que aún no conocemos. Debajo de cada partícula se da su masa en una cierta unidad, llamada MeV. Un MeV es igual a 1.783×10^{-27} gramos. La razón por la cual se escogió esta unidad es histórica y no entraremos en esto ahora. Las partículas más pesadas se colocan más arriba en el cuadro; vemos que un neutrón y un protón tienen casi misma masa. En las columnas verticales hemos colocado las partículas con la misma carga eléctrica, todos los objetos neutros en una columna, todos los cargados positivamente, a la derecha de ésta, y todos los objetos cargados negativamente a la izquierda.

Las partículas se indican con una línea continua y las "resonancias" con una a trazos. Varias partículas se han omitido en la tabla. Éstas incluyen las importantes partículas de masa cero y carga cero, el fotón y el gravitón, que no caen en el esquema de clasificación de bariones-mesones-leptones y también algunas de las resonancias más nuevas (K*, ϕ, η). Las antipartículas de los mesones están indicadas en la tabla, pero las antipartículas de los leptones y bariones deberían indicarse en otra tabla, que sería exactamente igual a ésta reflejada en la columna cero. A pesar de que todas las partículas, excepto el electrón, el neutrino, el fotón, el gravitón y el protón son inestables, los productos de desintegración se han indicado sólo para las resonancias. Asignaciones de extrañeza no son aplicables para leptones, dado que éstos no interactúan fuertemente con los núcleos.

Todas las partículas que están junto con los neutrones y protones se llaman *bariones*, y existen las siguientes: hay una "lambda", con una masa de 1 115 MeV, y otras tres llamadas sigmas (menos, cero y más), con diversas masas casi iguales. Hay grupos de multipletes con casi la misma masa dentro del uno o dos por ciento. Cada partícula de un multiplete tiene la misma extrañeza. El primer multiplete es el doblete protón-neutrón y enseguida hay un singlete (la lambda), después el triplete sigma y finalmente el doblete xi. Muy recientemente, en 1961, se han encontrado algunas partículas más. Pero, *¿son realmente* partículas? Viven un tiempo tan corto, se desintegran casi instantáneamente tan pronto como se han formado, que no sabemos si se pueden considerar como nuevas partículas o cierta especie de interacción de "resonancia" de una cierta energía definida entre los productos Λ y π en los cuales ellas se desintegran.

Además de los bariones, las otras partículas que intervienen en la interacción nuclear se llaman *mesones*. Hay primero los piones, que se presentan en tres variedades, positivo, negativo y neutro; forman otro multiplete. Hemos encontrado también algunas cosas nuevas, llamadas mesones K, y se presentan como un doblete, K^+ y K^0. También, cada partícula tiene su antipartícula, salvo que una partícula *sea su propia* antipartícula. Por ejemplo, el π^- y el π^+ son antipartículas, pero el π^0 es su propia antipartícula. Son antipartículas el K^- y el K^+, y el K^0 y el $\overline{K}^0$. Además, en 1961 encontramos también algunos mesones más, o *quizá* mesones que se desintegran casi inmediatamente. Una cosa llamada ω, que se transforma en tres piones, tiene una masa de 780 en esta escala y algo menos seguro es un objeto que se desintegra en dos piones. Estas partículas, llamadas mesones y bariones, y las antipartículas de los mesones están en el mismo cuadro, pero las antipartículas de los bariones deben colocarse en otro cuadro, "reflejado" en la columna de carga cero.

Del mismo modo que la tabla de Mendeléyev era muy buena, excepto por el hecho de que existía un número de elementos, llamados tierras raras, que quedaban colgando sueltos fuera de ella; aquí tenemos una cantidad de cosas que cuelgan sueltas fuera de esta tabla —partículas que no interactúan fuertemente en los núcleos, no tienen nada que ver con una interacción nuclear y no tienen una interacción fuerte (me refiero al poderoso tipo de interacción de energía nuclear)—. Se llaman leptones y son los siguientes: está el electrón que tiene una masa muy pequeña en esta escala, sólo 0.510 MeV. Está después este otro, el muon que tiene una masa mucho mayor, 206 veces más pesado que un electrón. Hasta donde podemos decir, de acuerdo con todos los experimentos hechos hasta ahora, la diferencia entre el electrón y el muon no es más que la masa. Todo se verifica exactamente igual para el muon que para el electrón, excepto que uno es más pesado que el otro. ¿Por qué existe otro más pesado? ¿Para qué sirve? No lo sabemos. Además existe un leptón que es neutro, llamado neutrino, y esta partícula tiene masa cero. En realidad, se sabe ahora que hay *dos* tipos diferentes de neutrinos, uno relacionado con los electrones y el otro relacionado con los muones.

Finalmente, tenemos otras dos partículas que no interactúan fuertemente con las nucleares: una es el fotón, y quizá, si el campo gravitacional también tiene un análogo cuántico (hasta ahora no se ha elaborado una teoría cuántica de la gravitación), entonces existiría una partícula, un gravitón, que tendría masa cero.

¿Qué es esta “masa cero”? Las masas dadas aquí son las masas de las partículas *en reposo*. El hecho de que una partícula tenga masa cero significa, de cierto modo, que no puede *estar en reposo*. Un fotón nunca está en reposo, siempre se mueve a 300 000 kilómetros por segundo. Entenderemos más lo que significa masa, cuando comprendamos la teoría de la relatividad, que será introducida a su debido tiempo.

Estamos así frente a un gran número de partículas que parecen ser en conjunto los constituyentes fundamentales de la materia. Afortunadamente estas partículas no son *todas* diferentes en sus *interacciones* mutuas. En realidad parece haber precisamente *cuatro tipos* de interacción entre partículas, las cuales en orden de intensidad decreciente son la fuerza nuclear, las interacciones eléctricas, la interacción de desintegración beta y la gravedad. El fotón está acoplado a todas las partículas cargadas y la intensidad de la interacción está medida por un cierto número, que es $1/137$. La ley detallada de este acoplamiento se conoce, constituye la electrodinámica cuántica. La gravedad está acoplada a toda la *energía*, pero su acoplamiento es extremadamente débil, mucho más débil que el de la electricidad. Esta ley también se conoce. Después están las así llamadas desintegraciones débiles —la desintegración beta, que hace que un neutrón se desintegre en forma relativamente lenta en protón, electrón y neutrino—. Esta ley se conoce sólo parcialmente. La así llamada interacción fuerte, la interacción mesón-barión, tiene una intensidad 1 en esta escala y la ley se desconoce completamente, a pesar de que hay muchas reglas conocidas, como la de que el número de bariones no cambia en ninguna reacción.

Ésta es entonces la horrible condición de nuestra física actual. Para resumirla, yo diría esto: fuera del núcleo parece que sabemos todo: dentro de él, es válida la

Tabla 2-3. Interacciones elementales

Acoplamiento	Intensidad*	Ley
Fotón a partículas cargadas	$\sim 10^{-2}$	Ley conocida
Gravedad a toda energía	$\sim 10^{-40}$	Ley conocida
Desintegraciones débiles	$\sim 10^{-5}$	Ley parcialmente conocida
Mesones a bariones	~ 1	Ley desconocida (algunas reglas conocidas)

* La “intensidad” es una medida dimensional de la constante de acoplamiento que interviene en toda interacción ($\sim$ significa “del orden”).

mecánica cuántica —no se ha encontrado que fallen los principios de la mecánica cuántica—. El escenario donde ponemos todo nuestro conocimiento, diríamos que es el espacio-tiempo relativístico; quizá la gravedad esté contenida en el espacio-tiempo. No sabemos cómo se inició el universo y nunca hemos hecho experimentos que comprueben precisamente nuestras ideas del espacio y del tiempo por debajo de una distancia muy pequeña, por lo que así *sabemos* solamente que nuestras ideas funcionan por encima de esta distancia. Deberíamos también agregar que las reglas del juego son los principios de la mecánica cuántica y esos principios se aplican, hasta donde podemos decir, tanto a las partículas nuevas como a las antiguas. El origen de las fuerzas en los núcleos nos conduce a nuevas partículas, pero desgraciadamente aparecen en gran profusión y nos falta una comprensión completa de sus interrelaciones, a pesar de que ya sabemos que existen algunas relaciones muy sorprendentes entre ellas. Parece que estuviéramos tanteando gradualmente hacia una comprensión del mundo de las partículas subatómicas, pero en realidad no sabemos hasta dónde debemos ir todavía en esta faena.

3

La relación de la física con otras ciencias

3-1 Introducción

La física es la más fundamental y general de las ciencias, y ha tenido un profundo efecto en todo el desarrollo científico. En realidad, la física es el equivalente actual de lo que solía llamarse *filosofía natural*, de la cual provienen la mayoría de nuestras ciencias modernas. Estudiantes de muchas disciplinas se encuentran estudiando física a causa del rol básico que ésta juega en todos los fenómenos. En este capítulo trataremos de explicar cuáles son los problemas fundamentales en las otras ciencias, pero, por supuesto, es imposible realmente tratar en un espacio tan reducido las materias complejas, sutiles y hermosas de esos otros campos. La falta de espacio también impide que discutamos la relación entre la física y la ingeniería, la industria, la sociedad y la guerra, o aun la más notable relación entre la matemática y la física. (La matemática no es una ciencia desde nuestro punto de vista, en el sentido que no es una ciencia *natural*. El experimento no es una prueba de su validez). Debemos, incidentalmente, dejar en claro desde un comienzo que si una cosa no es una ciencia, no es necesariamente mala. Por ejemplo, el amor no es una ciencia. De modo que, si se dice que algo no es una ciencia, no significa que haya algo malo en esto; significa simplemente que no es una ciencia.

3-2 Química

La ciencia que es quizás la más profundamente afectada por la física es la química. Históricamente, en su comienzo, la química trataba casi enteramente de lo que ahora llamamos química inorgánica, la química de las sustancias que no están asociadas con los objetos vivientes. Se necesitó de un análisis considerable para descubrir la existencia de muchos elementos y sus relaciones —cómo forman los numerosos compuestos relativamente simples que se encuentran en las rocas, la tierra, etc.—. Esta química primitiva fue muy importante para la física. La interacción entre las dos ciencias era muy grande porque la teoría de los átomos fue comprobada en gran parte con experimentos de química. La teoría de la química, es decir, de las reacciones mismas, fue resumida ampliamente en la tabla periódica de Mendeléyev, la cual establece numerosas relaciones extrañas entre los diversos elementos, y fue la colección de reglas sobre qué sustancia se combina con cuál otra y cómo, lo que constituyó la química inorgánica. Todas estas reglas fueron últimamente explicadas en principio por la mecánica cuántica, de manera que la química teórica es en realidad física. Por otro lado, debe ponerse énfasis en que esta explicación es *en principio*. Ya hemos discutido la diferencia entre saber las reglas del juego de ajedrez y ser capaz de jugar. De manera que podemos conocer las reglas, pero no podemos jugar muy bien. Resulta así muy difícil predecir precisamente qué sucederá en una reacción química dada; sin embargo, la parte más profunda de la química teórica debe terminar en la mecánica cuántica.

Hay también una rama de la física y la química que ambas ciencias desarrollaron conjuntamente y que es extremadamente importante. Éste es el método estadístico aplicado a una situación en que hay leyes mecánicas, que se llama, propiamente, *mecánica estadística*. En cualquier situación química está implicado un gran número de átomos y hemos visto que los átomos se agitan todos en una forma complicada y casual. Si pudiéramos analizar cada colisión y fuéramos capaces de seguir en detalle el movimiento de cada molécula, esperaríamos poder deducir lo que sucede, pero los muchos números que se necesitan para seguir la

trayectoria de todas esas moléculas exceden tan enormemente la capacidad de cualquier computador, y ciertamente la capacidad de la mente, que fue necesario desarrollar un método para tratar con tales situaciones complicadas. La mecánica estadística es entonces la ciencia de los fenómenos del calor, o la termodinámica. La química inorgánica es una ciencia, ahora reducida esencialmente a lo que se llaman la físico-química y la química cuántica; la físico-química para estudiar las velocidades con que ocurren las reacciones y qué es lo que está sucediendo en detalle (¿cómo chocan las moléculas? ¿Cuáles partes saltan primero?, etcétera), y la química cuántica para ayudarnos a comprender lo que sucede en términos de las leyes físicas.

La otra rama de la química es la *química orgánica*, la química de las sustancias que están asociadas con las cosas vivientes. Por un tiempo se creyó que las sustancias que están asociadas con las cosas vivas eran tan maravillosas que no podían ser hechas a mano a partir de materiales inorgánicos. Esto no es absolutamente cierto; son exactamente lo mismo que las sustancias hechas en química inorgánica, pero comprenden disposiciones más complicadas de los átomos. Evidentemente, la química orgánica tiene una relación muy estrecha con la biología que suministra sus sustancias y con la industria; más aún, puede aplicarse mucha físico-química y mecánica cuántica tanto a los compuestos orgánicos como a los inorgánicos. Sin embargo, los principales problemas de la química orgánica no están en esos aspectos, sino más bien en el análisis y síntesis de las sustancias que se forman en los sistemas biológicos, en las cosas vivas. Esto conduce imperceptiblemente, a pasos, hacia la bioquímica y luego a la biología misma, o biología molecular.

3-3 Biología

Así llegamos a la ciencia de la *biología*, que es el estudio de las cosas vivas. En los primeros días de la biología, los biólogos tenían que tratar con problemas puramente descriptivos de buscar *qué* cosas vivas había, y así ellos tenían sólo que contar cosas como los pelos de los miembros de las pulgas. Después que estos asuntos fueron resueltos con gran interés, los biólogos se fueron hacia la *maquinaria* interior de los cuerpos vivos, primero desde un punto de vista global, naturalmente, porque se requiere algún esfuerzo para entrar en los detalles más finos.

Había una interesante relación primaria entre la física y la biología en la cual la biología ayudaba a la física en el descubrimiento de la *conservación de la energía*, lo cual fue primeramente demostrado por Mayer en conexión con la cantidad de calor que recibe y cede una criatura viva.

Si miramos más de cerca a los procesos biológicos de los animales vivos, vemos *muchos* fenómenos físicos: la circulación de la sangre, bombas, presión, etc. Hay nervios: sabemos qué es lo que pasa cuando nos paramos sobre una piedra puntiaguda, y de cierta manera u otra la información va desde la pierna hacia arriba. Es interesante cómo sucede. En sus estudios sobre los nervios, los biólogos han llegado a la conclusión que los nervios son tubos muy finos con una compleja pared, que es muy delgada; a través de esta pared la célula bombea iones; así que hay iones positivos en el exterior y negativos en el interior, como en un condensador eléctrico. Ahora bien, esta membrana tiene una propiedad interesante; si se "descarga" en un lugar, es decir, si algunos iones son capaces de atravesar en algún lugar de manera que allí se reduce el voltaje eléctrico, dicha influencia eléctrica se hace sentir sobre los iones vecinos y afecta la membrana de tal manera, que deja pasar también los iones en los puntos vecinos. Esto a su vez la afecta más allá, etc., y así hay una onda de "penetrabilidad" de la membrana que recorre la fibra cuando está "excitada" en un extremo al pararse sobre una piedra puntiaguda. Esta onda es algo análogo a una larga secuencia de fichas de dominó verticales; si se empuja la del extremo, ésta empuja a la próxima, etc. Por cierto, esto transmitirá solamente un mensaje, a no ser que las fichas de dominó se paren de nuevo; y análogamente en una célula nerviosa hay procesos que bombean lentamente de nuevo los iones hacia afuera para que el nervio quede

listo para el próximo impulso. Así es como sabemos lo que estamos haciendo (o por lo menos dónde estamos). Por supuesto, los efectos eléctricos asociados con este impulso nervioso pueden ser captados con instrumentos eléctricos y, debido a que *son* efectos eléctricos, evidentemente la física de los efectos eléctricos ha tenido mucha influencia en la comprensión del fenómeno.

El efecto opuesto es que, desde algún lugar del cerebro, se envía hacia afuera un mensaje a lo largo de un nervio. ¿Qué sucede en el extremo del nervio? Allí el nervio se ramifica en cositas finas, conectadas a una estructura cerca de un músculo, llamada placa terminal. Por razones que no son exactamente comprendidas, cuando un impulso llega al término del nervio, se eyectan pequeños paquetes (5 a 10 moléculas de una vez) de un compuesto químico llamado acetilcolina y ellos afectan la fibra muscular y la hacen contraerse —¡cuán simple! ¿Qué hace que se contraiga un músculo? Un músculo es un número muy grande de fibras muy cerca unas de otras, que contiene dos sustancias diferentes, miosina y actomiosina, pero el mecanismo mediante el cual la reacción química inducida por la acetilcolina puede modificar las dimensiones del músculo es aún desconocido—. Así, los procesos fundamentales en el músculo que producen los movimientos mecánicos no son conocidos.

La biología es un campo tan enormemente vasto que hay montones de otros problemas que ni siquiera podemos mencionar —problemas de cómo funciona la visión (qué produce la luz en el ojo), cómo funciona el oído, etc.—. (La forma en que funciona el *pensamiento* la discutiremos más tarde bajo la psicología). Bien, esas cosas concernientes a la biología que hemos discutido aquí no son, desde un punto de vista biológico, realmente fundamentales en el fondo de la vida, en el sentido de que incluso si las entendiéramos todavía no comprenderíamos la vida misma. Para dar un ejemplo: los hombres que estudian los nervios estiman que su trabajo es muy importante porque, después de todo, usted no puede tener animales sin nervios. Pero se *puede* tener *vida* sin nervios. Las plantas no tienen ni nervios ni músculos, pero están funcionando, están igualmente vivas. Así, para los problemas fundamentales de la biología debemos observar más profundamente; cuando así hacemos, descubrimos que todos los seres vivientes tienen un gran número de características en común. El rasgo más común es que están hechos de *células*, dentro de cada una de las cuales hay un mecanismo complejo para hacer cosas químicamente. En las células de las plantas, por ejemplo, hay un mecanismo para recoger luz y generar glucosa, la que es consumida en la oscuridad para mantener la planta viva. Cuando la planta es ingerida, la glucosa genera en el animal una serie de reacciones químicas muy estrechamente relacionadas con la fotosíntesis (y su efecto opuesto en la oscuridad) en las plantas.

En las células de los sistemas vivos hay muchas reacciones químicas complicadas en las cuales un compuesto se transforma en otro y otro. Para dar una idea del enorme esfuerzo que se ha hecho en el estudio de la química, la figura 3-1 resume nuestro conocimiento actualizado sobre justamente una pequeña parte de las muchas series de reacciones que ocurren en las células, quizá un pequeño porcentaje o algo así de ellas.

Aquí vemos una serie entera de moléculas que cambian de una a otra en una secuencia o ciclo de pasos más bien pequeños. Se le llama el ciclo de Krebs, el ciclo respiratorio. Cada uno de los compuestos químicos y cada uno de los pasos es bastante simple, en función de qué cambios se hacen en la molécula, pero —y esto es un descubrimiento central importante en bioquímica— estos cambios son *relativamente difíciles de llevar a cabo en un laboratorio.* Si tenemos una sustancia y otra muy similar, la primera no se convierte simplemente en la otra porque las dos formas están corrientemente separadas por una barrera o "loma" de energía. Consideren esta analogía: si queremos trasladar un objeto de un lugar a otro que está en el mismo nivel pero en el otro lado de una loma, podemos empujarlo por encima de la cumbre; pero hacerlo así requiere que se le agregue alguna energía. Así la mayoría de las reacciones químicas no ocurren, porque hay lo que se llama una *energía de activación* de por medio. Para agregar un átomo extra a nuestro compuesto químico se necesita que lo *acerquemos* lo suficiente para que pueda ocurrir un reordenamiento; entonces se pegará. Pero si no podemos darle

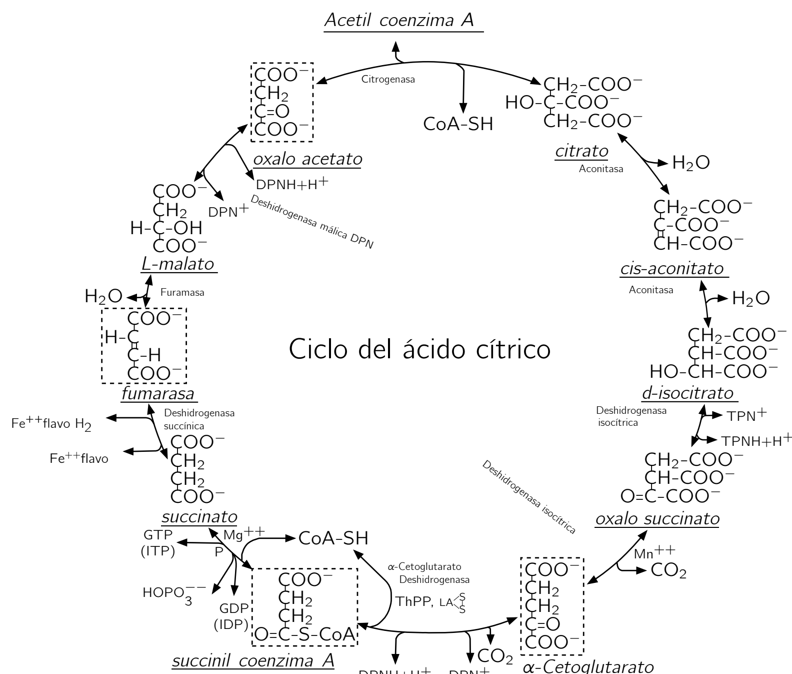


Figura 3-1. El ciclo de Krebs.

suficiente energía para acercarlo suficientemente, no completará el propósito, realizará parte del camino hacia arriba de la loma y volverá hacia abajo de nuevo. Sin embargo, si pudiéramos literalmente tomar las moléculas en nuestras manos y empujar y tirar los átomos alrededor de tal manera que se abriera un hueco para permitir la entrada de un nuevo átomo, y luego dejarlos volver, habríamos encontrado otro camino *alrededor* de la loma, el cual no necesitaría de energía extra, y la reacción procedería fácilmente. Ahora, realmente *hay* en las células moléculas *muy* grandes, mucho más grandes que aquellas cuyos cambios hemos estado describiendo, que en alguna forma complicada sujetan a las moléculas pequeñas en forma adecuada para que la reacción pueda realizarse fácilmente. Esas cosas muy grandes y complicadas se llaman *enzimas*. (Primeramente se llamaron fermentos porque se descubrieron originalmente en la fermentación del azúcar. En realidad, algunas de las primeras reacciones en el ciclo fueron descubiertas allí). La reacción procederá en presencia de una enzima.

Una enzima está hecha de otra sustancia llamada *proteína*. Las enzimas son muy complicadas y cada una es diferente, estando cada una construida para controlar cierta reacción especial. Los nombres de las enzimas están escritos en la figura 3-1 en cada reacción. (Algunas veces la misma enzima puede controlar dos reacciones.) Ponemos énfasis en que las enzimas mismas no intervienen directamente en la reacción. Ellas no cambian; solamente dejan pasar un átomo de un lugar a otro. Habiendo hecho esto, la enzima está lista para hacerlo con la próxima molécula, como la máquina de una fábrica. Por cierto, debe haber un suministro de ciertos átomos y una forma de disponer de otros átomos. Tómese el hidrógeno, por ejemplo: hay enzimas que sobre ellas tienen unidades especiales que transportan el hidrógeno para todas las reacciones químicas. Por ejemplo, hay tres o cuatro enzimas hidrógeno-reductoras que se usan en diferentes lugares sobre todo nuestro ciclo. Es interesante que el mecanismo que libera algún hidrógeno en un sector tomará aquel hidrógeno y lo empleará en algún otro lugar.

La característica más importante del ciclo de la figura 3-1 es la transformación de GDP en GTP (guanosín difosfato en guanosín trifosfato) porque una sustancia contiene mucha más energía que la otra. Justamente, así como hay una "caja" en ciertas enzimas para transportar átomos de hidrógeno, hay ciertas "cajas" transportadoras de *energía* que implican al grupo trifosfato. Entonces, la GTP tiene más energía que la GDP y si el ciclo se desplaza en un sentido, estamos produciendo moléculas que tienen una energía extra que puede movilizar otro ciclo que *requiere* energía, por ejemplo la contracción de un músculo. El músculo

no se contraerá a no ser que haya GTP. Podemos tomar fibra muscular, ponerla en agua y agregar GTP, y las fibras se contraerán transformando GTP en GDP si está presente la enzima correcta. Así, el sistema real está en la transformación GDP-GTP; en la oscuridad la GTP, que se ha estado almacenando durante el día, se usa para producir el ciclo completo en la otra dirección. Una enzima, observen, no se preocupa en qué dirección procede la reacción pues, si lo hiciera, violaría una de las leyes de la física.

La física es de gran importancia en la biología y otras ciencias por otra razón que tiene que ver con *técnicas experimentales*. En realidad, si no fuera por el gran desarrollo de la física experimental, estos cuadros bioquímicos no se conocerían hoy. La razón es que la herramienta más útil de todas para el análisis de este sistema fantásticamente complejo es *marcar* los átomos que se usan en las reacciones. Así, si pudiéramos introducir en el ciclo algún dióxido de carbono que tiene una "marca verde", y luego medir después de tres segundos donde está la marca verde, y de nuevo medir después de 10 segundos, etc., podríamos seguir el curso de las reacciones. ¿Qué son las "marcas verdes"? Son *isótopos* diferentes. Recordemos que las propiedades químicas de los átomos están determinadas por el número de *electrones*, no por la masa del núcleo. Pero puede haber, en el carbono por ejemplo, seis o siete neutrones junto con los seis protones que tienen todos los núcleos de carbono. Químicamente, los dos átomos C^{12} y C^{13} son iguales, pero difieren en peso y tienen propiedades nucleares diferentes, y por eso son distinguibles. Usando estos isótopos de pesos diferentes, o aun isótopos radioactivos como el C^{14}, lo que da un medio más sensible para seguir cantidades muy pequeñas, es posible seguir las reacciones.

Ahora volvemos a la descripción de enzimas y proteínas. No todas las proteínas son enzimas, pero todas las enzimas son proteínas. Hay muchas proteínas, como las proteínas de los músculos, las proteínas estructurales que están, por ejemplo, en los cartílagos, pelo, piel, etc., que no son enzimas en sí mismas. Sin embargo, las proteínas son una sustancia muy característica de la vida: en primer lugar forman todas las enzimas, y en segundo forman gran parte del resto de la materia viviente. Las proteínas tienen una estructura muy interesante y simple. Son una serie, o cadena, de diferentes *aminoácidos*. Hay 20 aminoácidos diferentes, y todos ellos pueden combinarse entre sí para formar cadenas cuya espina dorsal es CO-NH, etc. Las proteínas no son otra cosa que cadenas de varios de estos 20 aminoácidos. Cada uno de los aminoácidos sirve probablemente para algún propósito especial. Algunos, por ejemplo, tienen un átomo de azufre en cierto lugar; cuando dos átomos de azufre están en la misma proteína, forman un enlace, esto es, unen la cadena en dos puntos y forman un anillo. Otro tiene átomos de oxígeno extra que lo hace una sustancia ácida, otro tiene una característica básica. Algunos tienen grandes grupos colgando hacia afuera por un lado, de modo que ocupan mucho espacio. Uno de los aminoácidos llamado prolina no es realmente un aminoácido, sino un aminoácido. Hay una pequeña diferencia: cuando la prolina está en la cadena hay un retorcimiento en ésta. Si quisiéramos producir una proteína en particular, daríamos estas instrucciones: ponga uno de esos ganchos de azufre aquí; luego agregue algo que ocupe lugar; ahora prenda algo para poner una enroscadura en la cadena. En esta forma obtendremos una cadena de aspecto complicado, enganchada a sí misma y teniendo una estructura compleja; presumiblemente ésta es justamente la manera en que se forman todas las variedades de enzimas. Uno de los grandes triunfos en los tiempos recientes (desde 1960) fue descubrir por fin la disposición atómica espacial exacta de ciertas proteínas, las que envuelven unos 56 a 60 aminoácidos en hilera. Más de un millar de átomos (más cercano a 2 000, si contamos los átomos de hidrógeno) han sido localizados en una estructura compleja en dos proteínas. La primera fue la hemoglobina. Uno de los aspectos tristes de este descubrimiento es que no podemos ver cosa alguna de esta estructura; no comprendemos cómo funciona y la forma en que lo hace. Por cierto, ése es el próximo problema por atacar.

Otro problema es: ¿cómo saben las enzimas qué cosa son? Una mosca de ojos rojos forma una mosca de ojos rojos, y así la información de la estructura entera de enzimas para hacer un pigmento rojo debe pasar de una mosca a la siguiente.

Esto es hecho por una sustancia en el núcleo de la célula, no una proteína, llamada adn's (abreviación de ácido desoxirribonucleico). Ésta es la sustancia clave que pasa de una célula a otra (por ejemplo, las células de semen consisten principalmente en ADN) y lleva la información de cómo hacer las enzimas. El ADN es el "programa". ¿Qué aspecto tiene el programa y cómo funciona? Primero, el programa debe ser capaz de reproducirse a sí mismo. Segundo, debe ser capaz de instruir a las proteínas. Respecto a la reproducción, podemos pensar que procede como en la reproducción celular. Las células simplemente agrandan su tamaño y luego se dividen por la mitad. ¿Deberá ser así con las moléculas de ADN, entonces, de modo que ellas también aumentan de tamaño y se dividen por la mitad? ¡Cada *átomo* por cierto no aumenta de tamaño y se divide por la mitad! No, es imposible reproducir una molécula excepto de un modo algo más inteligente.

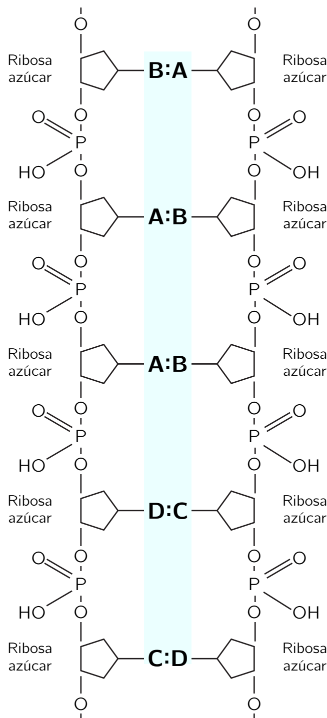


Figura 3-2. Diagrama esquemático del ADN.

La estructura de la sustancia del ADN fue estudiada por un largo tiempo, primero químicamente para encontrar la composición, y luego con rayos X para encontrar la estructura en el espacio. El resultado fue el siguiente descubrimiento notable: la molécula de ADN es un par de cadenas enrolladas una sobre la otra. La espina dorsal de cada una de estas cadenas, que son análogas a las cadenas de las proteínas pero químicamente bastante diferentes, es una serie de grupos azúcar y fosfato, como se muestra en la figura 3-2. Ahora vemos cómo la cadena puede contener instrucciones, pues si pudiéramos dividir esta cadena por la mitad, tendríamos una serie $BAADC\ldots$ y cada cosa viviente podría tener una serie diferente. En consecuencia, tal vez, de alguna manera, las *instrucciones* específicas para la manufactura de proteínas están contenidas en la *serie* específica del ADN.

Unidos a cada azúcar a lo largo de la línea y ligando las dos cadenas entre sí hay ciertos pares de uniones cruzadas. Sin embargo, no son todas de la misma naturaleza; hay cuatro tipos, llamados adenina, timina, citosina y guanina, pero llamémoslas A, B, C y D. La cosa interesante es que sólo ciertos pares pueden situarse uno opuesto al otro, por ejemplo A con B y C con D. Estos pares están puestos en ambas cadenas de tal manera que "encajan entre sí" y tienen una fuerte energía de interacción. Sin embargo, C no encajará con A, y B no encajará con C; sólo encajarán en pares, A contra B, y C contra D. Por lo tanto, si uno es C, el otro debe ser D, etc. Cualesquiera que sean las letras en una cadena, cada una debe tener una letra complementaria específica en la otra cadena.

¿Qué pasa entonces con la reproducción? Supóngase que separamos esta cadena en dos. ¿Cómo podemos hacer otra justamente como ésta? Si en las sustancias de las células hay un departamento productor que entrega fosfato, azúcar y unidades de A, B, C, D que no están conectadas en cadena, los únicos que se unirán a nuestra cadena separada serán los correctos, los complementos de $BAADC\ldots$, es decir, $ABBCD\ldots$ Así, lo que sucede es que la cadena se divide por la mitad durante la división celular, una mitad que terminará por ir con una célula y la otra mitad que terminará en la otra célula; cuando se separan, se forma una nueva cadena complementaria de cada mitad de cadena.

A continuación viene la pregunta, precisamente, ¿cómo determina el orden de las unidades A, B, C, D la disposición de los aminoácidos en la proteína? Éste es el problema central no resuelto actualmente en biología. Las primeras claves o partes de información, sin embargo, son éstas: hay en la célula pequeñas partículas llamadas ribosomas, y se sabe ahora que éste es el lugar donde se fabrican las proteínas. Pero los ribosomas no están en el núcleo, donde están el ADN y sus instrucciones. Algo parece que sucede. Sin embargo, también se sabe que pequeñas piezas moleculares se desprenden del ADN —no tan largas como la gran molécula de ADN que transporta toda la información en sí misma, pero como una pequeña sección de ella—. Ésta se llama ARN, pero esto no es esencial. Es una especie de copia del ADN, una copia resumida. El ARN, que en cierta forma lleva un mensaje acerca de qué proteína debe fabricar, va hacia el ribosoma. Eso es conocido. Tan pronto llega allí, la proteína es sintetizada en el ribosoma. Esto también es conocido. Sin embargo, los detalles de cómo los aminoácidos entran y se disponen de acuerdo con el código que está contenido

en el ARN son, por de pronto, aún desconocidos. No sabemos cómo leerlo. Si conociéramos, por ejemplo, la “alineación” A, B, C, C, A, no podríamos decirle a usted qué proteína será producida.

Ciertamente ninguna otra materia o campo está haciendo más progresos en tantos nuevos frentes en el momento presente que la biología, y si tuviéramos que nombrar la suposición más poderosa de todas que conducen a uno más y más a un intento de comprender la vida, es aquella en que *todas las cosas están hechas de átomos*, y en que todo lo que las cosas vivas hacen puede ser comprendido en términos de las agitaciones y movimientos de los átomos.

3-4 Astronomía

En esta explicación fugaz del mundo entero, debemos ahora pasar a la astronomía. La astronomía es más antigua que la física. En realidad, dio origen a la física al mostrar la hermosa simplicidad del movimiento de las estrellas y planetas, cuya comprensión fue el *comienzo* de la física. Pero el descubrimiento más notable de toda la astronomía es que *las estrellas están hechas de átomos de la misma naturaleza que los que se encuentran en la Tierra.** ¿Cómo se hizo esto? Los átomos liberan luz que tiene frecuencias definidas, algo así como el timbre de un instrumento musical que tiene tonos definidos o frecuencias de sonido. Cada vez que estamos escuchando varios tonos diferentes podemos distinguirlos, pero cuando miramos con nuestros ojos una mezcla de colores no podemos distinguir las partes de que está hecha, porque el ojo no es ni cercanamente tan discernidor como el oído a este respecto. Sin embargo, con un espectroscopio *podemos* analizar las frecuencias de las ondas luminosas y de esa manera podemos ver los tonos de los átomos que hay en las diferentes estrellas. De hecho, dos elementos químicos se descubrieron en una estrella antes que se descubrieran en la Tierra. El helio fue descubierto en el Sol, de ahí su nombre, y el tecnecio fue descubierto en ciertas estrellas frías. Esto, por supuesto, nos permite progresar en el entendimiento de las estrellas, porque ellas están hechas de los mismos tipos de átomos que existen en la Tierra. Ahora sabemos mucho acerca de los átomos, especialmente en cuanto a su comportamiento bajo condiciones de alta temperatura, pero no de alta densidad; así podemos analizar con la mecánica estadística el comportamiento de la sustancia estelar. Aun cuando no podemos reproducir las condiciones en la Tierra usando leyes básicas de la física, podemos a menudo decir precisamente, o muy aproximadamente, qué sucederá. Así es como la física ayuda a la astronomía. Por extraño que parezca, comprendemos la distribución de materia en el interior del Sol mucho mejor que lo que comprendemos el interior de la Tierra. Lo que sucede en el *interior* de una estrella se comprende mejor que lo que pudiera adivinarse de la dificultad de tener que mirar un pequeño punto luminoso a través de un telescopio, porque podemos *calcular* qué deben hacer los átomos en las estrellas en la mayoría de las circunstancias.

Uno de los descubrimientos más impresionantes fue el origen de la energía de las estrellas, que las hace continuar quemándose. Uno de los hombres que descubrió esto había salido con su amiga la noche siguiente de haberse dado cuenta de que en las estrellas se debían estar produciendo *reacciones nucleares*

* ¡Qué manera de precipitarme a través de esto! ¡Cuánto contenido tiene cada frase de esta breve historia! “Las estrellas están hechas de los mismos átomos que los de la Tierra.” Corrientemente yo tomo un pequeño tópico como éste para dictar una clase. Los poetas dicen que la ciencia elimina la belleza de las estrellas —meros globos de átomos de gas—. Nada es “mero”. Yo también puedo ver las estrellas en una noche despejada y sentirlas. ¿Pero veo yo más o menos? La vastedad de los cielos ensancha mi imaginación —clavado en este carrusel, mi pequeño ojo puede coger luz de un millón de años de edad—. Una vasta estructura de la cual yo soy una parte —quizá mi material fue arrojado de alguna estrella olvidada, como el que está arrojando una allí—. O verlas con el ojo más grande de Palomar, apartándose desde un punto común de partida donde quizás estuvieron todas juntas. ¿Cuál es la estructura, o el significado, o el *porqué*? No le hace daño al misterio conocer un poco de él. ¡Porque mucho más maravillosa es la verdad que la que cualquier artista en el pasado imaginó! ¿Por qué los poetas del presente no hablan de ella? ¿Qué hombres son los poetas que pueden hablar de Júpiter como si fuera un hombre, pero si es una inmensa esfera rotante de metano y amoniaco deben permanecer mudos?

para que brillaran. Ella dijo: "¡Mira qué bonito brillan las estrellas!" Él respondió: "Sí, y justamente ahora yo soy el único hombre en el mundo que sabe *por qué* brillan". Ella simplemente se rió de él. Ella no estaba impresionada de haber salido con el único hombre del mundo que, en ese momento, sabía por qué brillan las estrellas. Bueno, es triste estar solo, pero así son las cosas en este mundo.

Es la "combustión" nuclear del hidrógeno la que suministra la energía del Sol; el hidrógeno se convierte en helio. Además, en última instancia, la producción de los diversos elementos químicos se verifica en los centros de las estrellas a partir del hidrógeno. El material del que estamos hechos *nosotros* fue "cocinado" una vez en una estrella y escupido hacia afuera. ¿Cómo lo sabemos? Porque hay una clave. La proporción de los diferentes isótopos (cuánto C^{12}, cuánto C^{13}, etc.) es algo que nunca cambia en las reacciones *químicas*, porque las reacciones químicas son tan idénticas para las dos. Las proporciones son puramente el resultado de reacciones *nucleares*. Observando las proporciones de los isótopos en el rescoldo frío y apagado en que estamos, podemos descubrir cómo fue el *horno* donde se formaron los materiales de que estamos hechos. Aquel horno fue como las estrellas, y así es muy probable que nuestros elementos fueron "hechos" en las estrellas y escupidos en las explosiones que llamamos novas y supernovas. La astronomía está tan cerca de la física que estudiaremos muchas cosas astronómicas a medida que prosigamos.

3-5 Geología

Ahora pasemos a lo que se llama *ciencias de la Tierra*, o *geología*. Primero la meteorología y el tiempo. Por cierto, los *instrumentos* de meteorología son instrumentos físicos, y el desarrollo de la física experimental hizo posible estos, como se explicó anteriormente. Sin embargo, la teoría de la meteorología nunca ha sido investigada satisfactoriamente por los físicos. "Bien", dirán ustedes "no hay otra cosa que aire, y conocemos las ecuaciones de los movimientos del aire". Sí, es cierto. "Así, si sabemos las condiciones del aire de hoy, ¿por qué no podemos calcular las condiciones del aire de mañana?" Primero, no sabemos *realmente* cuál es la condición de hoy, porque el aire está arremolinándose y dando vueltas por todas partes. Resulta ser muy susceptible y aun inestable. Si han visto alguna vez correr suavemente el agua sobre una represa y luego convertirse en un gran número de burbujas y gotas cuando cae, comprenderán lo que quiero decir con inestable. Ustedes conocen la condición del agua antes que traspase el vertedero; es perfectamente tranquila; pero en el momento que comienza a caer, ¿dónde empiezan las gotas? ¿Qué determina lo grande que van a ser los trozos y dónde estarán? Esto no se sabe, porque el agua es inestable. Aun una masa de aire moviéndose suavemente, al traspasar una montaña se convierte en complejos remolinos y torbellinos. En muchos campos encontramos esta situación de *flujo turbulento* que no podemos analizar actualmente. ¡Dejemos rápidamente el asunto del tiempo y discutamos sobre geología!

El asunto básico para la geología es: ¿qué hace que la Tierra sea lo que es? Los procesos más obvios están al frente de nuestros ojos, los procesos de erosión de los ríos, los vientos, etc. Es muy fácil comprenderlos, pero por cada poco de erosión hay algo más que está sucediendo. Las montañas no son más bajas hoy en promedio de lo que fueron en el pasado. Debe haber procesos *formadores* de montañas. Encontrarán, si estudian geología, que *hay* procedimientos formadores de montañas y volcanismos, los que nadie comprende pero que son la mitad de la geología. El fenómeno de los volcanes no se comprende realmente. Lo que produce un terremoto, a la postre, no se comprende. Se comprende que si hay algo empujando a otra cosa, cede repentinamente y se desliza, eso está bien. Pero ¿qué es lo que empuja y por qué? La teoría es que hay corrientes en el interior de la Tierra —corrientes circulantes, debido a la diferencia de temperatura interior y exterior—, las cuales en su movimiento empujan ligeramente la superficie. Así, si hay dos circulaciones opuestas vecinas, la materia se acumula en la región donde se juntan y forman cadenas de montañas que están en condiciones desafortunadas de tensión y así producen volcanes y terremotos.

¿Qué pasa en el interior de la Tierra? Mucho se sabe acerca de la velocidad de las ondas sísmicas a través de la Tierra y la distribución de densidades de la Tierra. Sin embargo, los físicos han sido incapaces de obtener una buena teoría sobre lo densa que deberá ser una sustancia a las presiones que se esperaría en el centro de la Tierra. En otras palabras, no podemos calcular las propiedades de la materia muy bien en dichas circunstancias. Lo hacemos mucho menos bien con la tierra que lo hacemos con las condiciones de la materia en las estrellas. La matemática implicada parece ser un poco difícil, hasta ahora, pero quizá no pasará mucho tiempo antes de que alguien se dé cuenta de que es un problema importante y que realmente lo solucione. El otro aspecto, por cierto, es que aun si supiéramos la densidad, no podríamos calcular las corrientes de circulación. Tampoco podemos realmente deducir las propiedades de las rocas a elevadas presiones. No podemos decir con qué rapidez las rocas "cederán"; todo eso debe ser resuelto por el experimento.

3-6 Psicología

Consideremos a continuación la ciencia de la *psicología.* Incidentalmente, el psicoanálisis no es una ciencia: en el mejor de los casos es un proceso médico, o quizá aún brujería. Tiene una teoría acerca de qué produce la enfermedad —muchos "espíritus" diferentes, etc.—. El hechicero tiene la teoría de que una enfermedad como la malaria está causada por un espíritu que viene del aire; no se sana agitando una culebra sobre él; en cambio, la quinina sí ayuda la malaria. Así, si están enfermos, yo les aconsejaría que vayan al hechicero, porque es el hombre en la tribu que sabe más acerca de enfermedades; por otro lado, su conocimiento no es ciencia. El psicoanálisis no ha sido cuidadosamente comprobado por el experimento, y no hay manera de encontrar una lista del número de casos en los cuales resulta, el número de casos en que no resulta, etcétera.

Las otras ramas de la psicología, que implican cosas como la fisiología de las sensaciones —qué sucede en el ojo, y qué sucede en el cerebro— son, si quieren, menos interesantes. Pero cierto progreso pequeño, pero real, se ha hecho al estudiarlas. Uno de los progresos técnicos más interesantes puede ser o no llamado psicología. El problema central de la mente, si quieren, o del sistema nervioso es éste: cuando un animal aprende algo, puede hacer algo diferente de lo que podía hacer antes y sus células cerebrales deben haber cambiado también, si es que está hecho de átomos. *¿En qué sentido es diferente?* No sabemos dónde mirar, o qué buscar, cuando algo se memoriza. No sabemos qué significa o qué cambio hay en el sistema nervioso cuando se aprende una realidad. Éste es un problema muy importante que no ha sido resuelto en absoluto. Suponiendo, sin embargo, que existe algo como la memoria, el cerebro es una masa tan enorme de alambres y nervios interconectados que probablemente no puede ser analizado en una forma directa. Hay una analogía de esto con las máquinas computadoras y los elementos de computación, en que también tienen muchas líneas, y que tienen algún tipo de elemento análogo, quizá, a la sinapsis o conexión de un nervio con otro. Esto es una materia muy interesante que no tenemos tiempo de discutir más allá —las relaciones entre el pensamiento y las máquinas computadoras—. Debe apreciarse, por cierto, que esta materia nos dirá muy poco acerca de las complejidades reales del comportamiento humano ordinario. Todos los seres humanos son tan diferentes. Pasará mucho tiempo antes que lleguemos ahí. Debemos empezar mucho más atrás. Si pudiéramos siquiera deducir cómo funciona un *perro*, habríamos avanzado bastante. Los perros son más fáciles de comprender, pero nadie sabe aún cómo funcionan estos animales.

3-7 ¿Cómo se llegó a eso?

Para que la física sea útil a las otras ciencias en una forma *teórica*, que no sea la invención de instrumentos, la ciencia en cuestión debe suministrar al físico una descripción del tema en el lenguaje del físico. Ellos pueden decir: "¿por qué salta una rana?, y el físico no puede contestar. Si ellos le dicen lo que es una

rana; que hay tantas moléculas; que hay un nervio aquí, etc., eso es diferente. Si ellos nos dijeran, más o menos, a qué se parecen la Tierra y las estrellas, entonces podemos resolverlo. Para que la teoría física sea de alguna utilidad, debemos saber exactamente dónde están colocados los átomos. Para comprender la química, debemos saber exactamente qué átomos están presentes, pues de lo contrario no podemos analizarla. Ésta es solamente la primera limitación, por supuesto.

Hay otro *tipo* de problema en las ciencias hermanas, que no existe en la física; podemos llamarlo, a falta de un término mejor, el asunto histórico. ¿Cómo se llegó a eso? Si comprendemos todo acerca de la biología, quisiéramos saber cómo todas las cosas que hay en la Tierra fueron a dar en ella. Existe la teoría de la evolución, una parte importante de la biología. En geología, no sólo queremos saber cómo se están formando las montañas, sino cómo se formó la Tierra entera en el comienzo, el origen del sistema solar, etc. Esto, por supuesto, nos conduce a querer saber qué tipo de materia existía en el mundo. ¿Cómo evolucionan las estrellas? ¿Cuáles fueron las condiciones iniciales? Éste es el problema de la historia de la astronomía. Se ha descubierto mucho acerca de la formación de las estrellas, de la formación de los elementos de los cuales estamos hechos y hasta un poco acerca del origen del universo.

No hay problemas históricos que se estén estudiando en la física actualmente. No tenemos una pregunta: "Aquí están las leyes de la física, ¿cómo se llegó a ellas?" No nos imaginamos, por ahora, que las leyes de la física están en cierto modo cambiando con el tiempo, que en el pasado fueran diferentes de lo que son en el presente. Por supuesto que *pueden* ser, y en el momento en que encontremos que *son*, la pregunta histórica de la física estará ligada con el resto de la historia del universo y entonces los físicos estarán hablando de los mismos problemas que los astrónomos, los geólogos y los biólogos.

Finalmente, hay un problema físico que es común a muchos campos, que es muy viejo y que no ha sido resuelto. No es el problema de encontrar nuevas partículas fundamentales, sino algo que quedó desde hace mucho tiempo atrás —más de 100 años—. Nadie en la física ha sido realmente capaz de analizarlo matemáticamente en forma satisfactoria a pesar de su importancia para las ciencias hermanas. Es el *análisis de fluidos circulantes* o *turbulentos*. Si observamos la evolución de una estrella, se llega a un punto donde podemos deducir que va a comenzar la convección, y a partir de esto ya no podemos deducir qué va a pasar. Unos pocos millones de años más tarde la estrella hace explosión, pero no podemos explicar la razón. No podemos analizar el tiempo. No conocemos los esquemas de los movimientos que debería haber en el interior de la tierra. La forma más simple del problema es tomar una cañería que es muy larga y empujar agua a través de ella a gran velocidad. Preguntamos: para empujar una cantidad dada de agua a través de esa cañería, ¿cuánta presión se necesita? Nadie puede analizarlo partiendo de principios primarios y de las propiedades del agua. Si el agua fluye muy lentamente, o si usamos algo espeso como la miel, entonces podemos hacerlo exactamente. Ustedes lo encontrarán en su texto. Lo que no podemos realmente hacer es tratar con agua real y fresca que corre a través de una cañería. Éste es el problema central que deberíamos resolver algún día y que no lo hemos hecho.

Decía una vez un poeta: "El universo entero está en un vaso de vino". Probablemente nunca sabremos lo que quería decir, pues los poetas no escriben para ser comprendidos. Pero es cierto que si miramos un vaso de vino lo suficientemente cerca, vemos el universo entero. Ahí están las cosas de la física: el líquido que se arremolina y se evapora dependiendo del viento y del tiempo, las reflexiones en el vidrio, y nuestra imaginación agrega los átomos. El vidrio es un destilado de las rocas terrestres y en su composición vemos los secretos de la edad del universo y la evolución de las estrellas. ¿Qué extraño arreglo de elementos químicos hay en el vino? ¿Cómo llegaron a ser? Están los fermentos, las enzimas, los sustratos y los productos. Allí en el vino se encuentra la gran generalización: toda vida es fermentación. Nadie puede descubrir la química del vino sin descubrir, como lo hizo Louis Pasteur, la causa de muchas enfermedades. ¡Cuán vívido es el vino tinto que imprime su existencia dentro del conocimiento de

quien lo observa! ¡Si nuestras pequeñas mentes, por alguna conveniencia, dividen este vaso de vino, este universo, en partes —física, biología, geología, astronomía, psicología, etc.—, recuerden que la naturaleza no lo sabe! Así, reunamos todo de nuevo sin olvidar en última instancia para qué sirve. Dejemos que nos dé un placer final más: ¡bébanlo y olvídense de todo!

4

Conservación de la energía

4-1 ¿Qué es la energía?

Habiendo terminado ya nuestra descripción general empezamos en este capítulo un estudio más detallado de los diferentes aspectos de la física. Para ilustrar las ideas y la clase de razonamiento que se puede usar en física teórica, examinaremos una de las leyes más básicas de la física: la conservación de la energía.

Hay un hecho, o si prefiere, una *ley*, que gobierna todos los fenómenos naturales conocidos hasta la fecha. No se conoce excepción a esta ley —es exacta hasta donde sabemos—. La ley se llama *conservación de la energía*. Establece que hay cierta cantidad que llamamos energía, que no cambia en los múltiples cambios que ocurren en la naturaleza. Ésta es una idea muy abstracta, porque es un principio matemático; significa que hay una cantidad numérica que no cambia cuando algo ocurre. No es la descripción de un mecanismo, o de algo concreto; ciertamente es un hecho raro que podamos calcular cierto número y que cuando terminemos de observar que la naturaleza haga sus trucos y calculemos el número otra vez, éste será el mismo. (Algo así como el alfil en un cuadro negro, que después de cierto número de movimientos —cuyos detalles son desconocidos— queda en el mismo cuadro. Es una ley de esta naturaleza). Puesto que ésta es una idea abstracta, ilustraremos su significado mediante una analogía.

Imaginemos un niño, tal vez "Daniel el Travieso", que tiene unos bloques que son absolutamente indestructibles, que no pueden dividirse en partes. Cada uno es igual al otro. Supongamos que tiene 28 bloques. Su madre lo coloca con los 28 bloques en una pieza al comenzar el día. Al *final* del día, por curiosidad, ella cuenta los bloques con mucho cuidado, y descubre una ley fenomenal —haga lo que haga con los bloques, ¡siempre quedan 28!—. Esto continúa por varios días, hasta que un día hay sólo 27 bloques, pero una pequeña investigación demuestra que hay uno bajo la alfombra —ella debe mirar por todas partes para estar segura de que el número de bloques no ha cambiado—. Un día, sin embargo, el número parece cambiar —hay sólo 26 bloques—. Una cuidadosa investigación indica que la ventana estaba abierta, y al mirar hacia afuera se encontraron los otros dos bloques. Otro día, una cuidadosa cuenta indica que ¡hay 30 bloques! Esto causa una gran consternación, hasta que se sabe que Bruce vino a visitarlo, trayendo sus bloques consigo y que dejó unos pocos en la casa de Daniel. Después de separar los bloques adicionales cierra la ventana, no deja entrar a Bruce, y entonces todo anda bien, hasta que una vez cuenta y encuentra sólo 25 bloques. Sin embargo, hay una caja en la pieza, una caja de juguetes, y la madre se dirige a abrir la caja de juguetes, pero el niño dice: "No, no abras mi caja de juguetes", y chilla. A la madre no le estaba permitido abrir la caja de juguetes. Como es extremadamente curiosa, y algo ingeniosa, inventa un ardid. Sabe que un bloque pesa 100 gramos, además pesó la caja cuando los 28 bloques están a la vista y pesó 500 gramos. Después, pesa la caja de nuevo, resta 500 gramos y divide entre 100. Ella descubre lo siguiente:

$$\begin{pmatrix}\text{número de}\\ \text{bloques vistos}\end{pmatrix} + \frac{(\text{peso de la caja}) - 500 \text{ gramos}}{100 \text{ gramos}} = \text{constante.} \tag{4.1}$$

Enseguida parece que hubiera algunas nuevas desviaciones, pero un estudio cuidadoso indica que el agua sucia de la bañera está cambiando de nivel. El niño está lanzando bloques al agua y ella no puede verlos porque está muy sucia, pero

puede saber cuántos bloques hay en el agua agregando otro término a su fórmula. Ya que la altura original del agua era de 15 centímetros y cada bloque eleva el agua medio centímetro, esta nueva fórmula sería:

$$\begin{aligned}\begin{pmatrix}\text{número de}\\ \text{bloques vistos}\end{pmatrix} &+ \frac{(\text{peso de la caja}) - 500 \text{ gramos}}{100 \text{ gramos}}\\ &+ \frac{(\text{altura del agua}) - 15 \text{ centímetros}}{0.5 \text{ centímetros}} = \text{constante}. \qquad (4.2)\end{aligned}$$

En el aumento gradual de la complejidad de su mundo, ella encuentra una serie completa de términos que representan modos de calcular cuántos bloques están en los lugares donde no le está permitido mirar. Como resultado, encuentra una fórmula compleja, una cantidad que *debe ser calculada*, que en esta situación siempre es la misma.

¿Cuál es la analogía de esto con la conservación de la energía? El más notable aspecto que debe ser abstraído de este cuadro es que *no hay bloques.* Quítese el primer término en (4.1) y en (4.2) y nos encontraremos calculando cosas más o menos abstractas. La analogía tiene los siguientes puntos. Primero, cuando estamos calculando la energía, a veces algo de ella deja el sistema y se va, y a veces algo entra. Para verificar la conservación de la energía debemos tener cuidado de no agregar ni quitar nada. Segundo, la energía tiene un número grande de *formas diferentes*, y hay una fórmula para cada tipo. Éstas son: energía gravitacional, energía cinética, energía térmica, energía elástica, energía eléctrica, energía química, energía radiante, energía nuclear y energía de masa. Si hacemos el total de las fórmulas para cada una de estas contribuciones, no cambiará a excepción de la energía que entra y que sale.

Es importante darse cuenta de que en la física actual no sabemos qué *es* la energía. No tenemos un modelo de energía formada por pequeñas gotas de un tamaño definido. No es así. Sin embargo, hay fórmulas para calcular cierta cantidad numérica, y cuando las juntamos todas nos da "28" —siempre el mismo número—. Es algo abstracto en el sentido de que no nos informa el mecanismo o las *razones* para las diversas fórmulas.

4-2 Energía potencial gravitacional

Puede entenderse la conservación de la energía sólo si tenemos la fórmula para todas sus formas. Deseo discutir la fórmula para la energía gravitacional cerca de la superficie de la Tierra, y deseo deducir esta fórmula de un modo que no tiene nada que ver con la historia, sino que es una simple línea de razonamiento inventada para esta lección en particular, esto a fin de darles a ustedes una ilustración del notable hecho que puede extraerse mucho acerca de la naturaleza a partir de unos pocos hechos y con un razonamiento acabado. Es una ilustración de la clase de trabajo que los físicos teóricos realizan habitualmente. Está modelado según el excelente argumento del señor Carnot sobre la eficiencia de las máquinas de vapor.*

Consideremos máquinas como una balanza —máquinas que tienen la propiedad de levantar un peso bajando otro—. Hagamos, además, una hipótesis: *que no existe movimiento perpetuo* para estas balanzas. (De hecho, no existe el movimiento perpetuo pues es un enunciado general de la ley de la conservación de la energía.) Debemos tener cuidado al definir el movimiento perpetuo. En primer lugar, hagámoslo para balanzas. Si cuando hemos levantado y bajado muchos pesos y llevado la máquina a su condición original, encontramos que el resultado neto es haber *levantado un peso*, entonces tenemos una máquina de movimiento perpetuo, porque podemos usar el peso levantado para poner en movimiento otra cosa. Es decir, *debe cumplirse* que la balanza después de levantar el peso vuelva a su exacta *condición original*, y además debe ser completamente *independiente*

* Aquí nuestro objetivo no es tanto el resultado (4.3), el cual de hecho ustedes ya pueden conocer, sino la posibilidad de llegar a él mediante un razonamiento teórico.

—esto es, que no haya recibido la energía de una fuente externa para levantar el peso— como los bloques de Bruce.

En la figura 4-1 se muestra una balanza muy simple para levantar pesos. Esta máquina "fuerte" levanta tres unidades de peso. Colocamos tres unidades en un platillo y una unidad en el otro. Sin embargo, a fin de hacerla trabajar realmente, debemos quitar un pequeño peso en el platillo de la izquierda. Por otra parte, podríamos levantar una unidad de peso bajando el peso de tres unidades, si trampeamos un poco quitando algo de peso del otro plato. Por supuesto, nos damos cuenta de que con cualquier balanza *real* debemos agregar una pequeña cantidad extra para lograr su funcionamiento. Esto no lo consideramos, *temporalmente*. Las máquinas ideales, aunque no existen, no necesitan nada extra. Una máquina que usemos en la realidad puede ser, por así decir, *casi* reversible: esto es, si levanta el peso de tres al bajar el de una, entonces también levantará aproximadamente el peso de una en la misma cantidad al bajar el peso de tres.

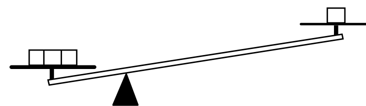


Figura 4-1. Una máquina simple para levantar pesos.

Imaginemos que hay dos clases de máquinas, las que *no* son reversibles, que incluyen todas las máquinas reales, y las que *son* reversibles, que, por supuesto no se consiguen en la realidad a pesar del cuidado que pongamos en el diseño de cojinetes, palancas, etc. Sin embargo, suponemos que existe una cosa tal —una máquina reversible— que baja una unidad de peso (un kilo o cualquier otra unidad) en una unidad de distancia y que al mismo tiempo levanta un peso de tres unidades. Llamemos Máquina A a esta máquina reversible. Supongamos que esta máquina reversible particular levante el peso de tres unidades una distancia X. A continuación, supongan que tenemos otra máquina, la Máquina B, que no es necesariamente reversible, que baja el peso de una unidad en una unidad de distancia, pero que levanta el peso de tres unidades una distancia Y. Podemos probar ahora que Y no es más grande que X; es decir, que es imposible construir una máquina que levante un peso *más alto* que una máquina reversible. Veamos por qué. Supongamos que Y fuera más grande que X. Tomemos un peso de una unidad y bajémoslo una unidad de altura con la máquina B, con lo que se eleva el peso de tres unidades una distancia Y. Entonces podríamos bajar el peso de Y a X, *obteniendo energía gratis*, y usar la máquina reversible A, funcionando a la inversa, para bajar el peso de tres unidades una distancia X y levantar el peso de una unidad a una unidad de altura. ¡Esto retornará el peso de una unidad a donde estaba antes y dejará ambas máquinas listas para ser usadas de nuevo! Por lo tanto, tendríamos movimiento perpetuo si Y fuera más grande que X, lo que asumimos que era imposible. Con esos supuestos debemos concluir que Y *no es más grande que* X, de modo que, de todas las máquinas que puedan diseñarse, la máquina reversible es la mejor.

Podemos ver, además, que todas las máquinas reversibles pueden levantar *exactamente la misma altura*. Supongan que B también fuera reversible. El razonamiento de que Y no es más alto que X es, por supuesto, tan bueno como antes, pero podemos hacer nuestro razonamiento de otra manera usando las máquinas en orden inverso y probar que X *no es más alto que* Y. Ésta es entonces una observación muy notable, ya que nos permite analizar la altura a la cual diferentes máquinas pueden levantar un objeto *sin mirar en el interior del mecanismo*. Sabemos de inmediato que si alguien construye una serie enormemente complicada de palancas que levantan tres unidades a una cierta distancia al bajar una unidad en una unidad de distancia, y las comparamos con una palanca simple que hace lo mismo y que es fundamentalmente reversible, su máquina no lo levantará más, sino tal vez menos. Si una máquina es reversible sabemos también exactamente *qué* altura alcanzará. Para resumir: cada máquina reversible, funcione como funcione, que un kilo baje un metro y levante un peso de tres kilos, siempre lo levantará la misma distancia X. Ésta es evidentemente una ley universal de gran utilidad. La pregunta siguiente es, por supuesto, ¿cuánto vale X?

Supongan que tenemos una máquina reversible que va a levantar una distancia X, tres unidades a una. Colocamos tres bolas en una estantería fija, como muestra la figura 4-2. Se mantiene una bola sobre una plataforma a una altura de un metro sobre el suelo. La máquina puede levantar tres bolas al bajar una a una distancia uno. Pues bien, hemos dispuesto que la plataforma que sostiene tres

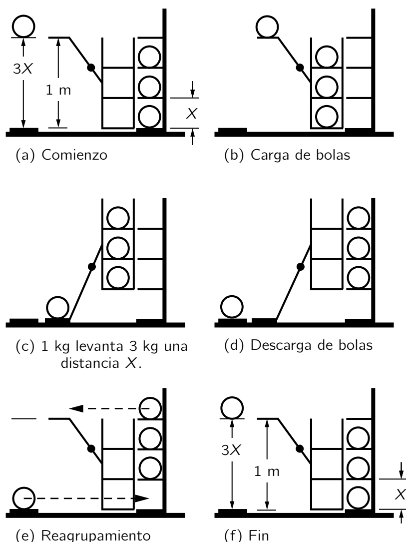


Figura 4-2. Una máquina reversible.

bolas tenga un piso y dos repisas, espaciadas exactamente a una distancia X, y además, que la estantería que sostiene las bolas están espaciadas a una distancia X, (a). Primero, hacemos rodar horizontalmente las bolas desde la estantería a las repisas, (b), y suponemos que esto no demanda energía, porque no cambiamos la altura. La máquina reversible opera entonces: baja la bola que está sola al suelo y levanta la plataforma una distancia X, (c). Más aún, hemos arreglado ingeniosamente la estantería de modo que estas bolas están de nuevo a la par con las repisas. Así descargamos las bolas a la estantería, (d); habiendo descargado las bolas podemos llevar la máquina a su condición original. Tenemos ahora tres bolas en las tres repisas superiores y una en el piso. Pero lo curioso es, por decirlo así, que no hemos levantado de ninguna manera *dos* de ellas, ya que después de todo, antes había bolas en las repisas dos y tres. El efecto resultante es haber levantado *una bola* la distancia $3X$. Pues bien, si $3X$ excede un metro, *podemos bajar* la bola para retornar la máquina a la condición inicial (f), y podemos hacer funcionar el aparato de nuevo. Por lo tanto, $3X$, no puede exceder un metro, porque si $3X$ excediera un metro podríamos realizar movimiento perpetuo. Asimismo, podemos probar que *un metro no puede exceder a* $3X$, cuando se hace funcionar la máquina al revés, ya que es una máquina reversible. Por lo tanto, $3X$ *no es ni mayor ni menor que un metro*, y entonces hemos descubierto, sólo con razonamientos, la ley de que $X = \frac{1}{3}$ de metro. La generalización es clara: un kilogramo cae cierta distancia al operar una máquina reversible; entonces la máquina puede levantar p kilogramos a esta distancia dividida por p. Otra forma de indicar este resultado es que el producto de tres kilogramos por la altura alcanzada, que en nuestro problema era X, es igual al producto de un kilogramo por la distancia que fue bajada, en este caso es un metro. Si tomamos todos los pesos y los multiplicamos por las alturas a las cuales están ahora por sobre el suelo, dejamos que la máquina funcione, e inmediatamente multiplicamos nuevamente todos los pesos por todas las alturas, *no habrá cambio.* (Tenemos que generalizar este ejemplo, en donde movemos sólo un peso, al caso en donde, cuando bajamos uno levantamos otros pesos —pero esto es fácil—.)

Llamamos *energía potencial gravitacional* la suma de los productos de los pesos por las alturas —la energía que tiene un objeto debido a su posición en el espacio con relación a la Tierra—. Entonces la fórmula para la energía gravitacional, siempre que no estemos muy lejos de la Tierra (la fuerza se debilita a medida que subimos), es:

$$\begin{pmatrix}\text{energía potencial}\\ \text{gravitacional}\\ \text{para un objeto}\end{pmatrix} = (\text{peso}) \times (\text{altura}). \tag{4.3}$$

Es una bellísima línea de razonamiento. El único problema es que tal vez no sea verdadera. (Después de todo, la naturaleza no *tiene* por qué marchar con nuestro razonamiento). Por ejemplo: quizás el movimiento perpetuo sea en efecto posible. Algunos de los supuestos pueden estar equivocados, o podemos haber cometido un error de razonamiento, de modo que siempre es necesario comprobar. En efecto, *resulta experimentalmente* cierto.

El nombre general para la energía que tiene que ver con la posición relativa a alguna otra cosa es energía *potencial.* Por supuesto, en este caso particular la llamamos *energía potencial gravitacional.* Si es cuestión de fuerzas eléctricas contra las que estamos trabajando, en vez de fuerzas gravitacionales, si estamos "levantando" cargas desde otras cargas, con muchas palancas, entonces el contenido de energía se llama *energía potencial eléctrica.* El principio general es que el cambio en la energía es el producto de la fuerza por la distancia que se desplaza la fuerza, y que esto es en general un cambio de la energía:

$$\begin{pmatrix}\text{cambio de}\\ \text{energía}\end{pmatrix} = (\text{fuerza}) \times \begin{pmatrix}\text{distancia en que}\\ \text{la fuerza actúa}\end{pmatrix}. \tag{4.4}$$

Volveremos a muchas de estas otras formas de energía a lo largo del curso.

El principio de la conservación de la energía es muy útil para deducir lo que ocurrirá en numerosas circunstancias. En la enseñanza media aprendimos muchas

leyes acerca de poleas y palancas usadas de diferentes maneras. Podemos ver ahora que estas “leyes” son *todas la misma cosa*, y que no teníamos necesidad de memorizar 75 reglas para darnos cuenta de ello. Un simple ejemplo es un plano inclinado liso, que por suerte es un triángulo cuyos lados son tres, cuatro y cinco unidades (figura 4-3). Colgamos el peso de un kilo sobre un plano inclinado con una polea, y un peso W al otro lado de la polea. Queremos saber cuánto debe pesar W para equilibrar un kilo en el plano. ¿Cómo podemos calcularlo? Si decimos que está justamente equilibrado, es reversible, y así puede moverse hacia arriba y hacia abajo, y podemos considerar la siguiente situación. En la situación inicial (a), el peso de un kilo está deabajo, y el peso W esta arriba. Cuando W se ha deslizado hacia abajo en forma reversible (b), tenemos el peso de un kilo arriba y el peso W a una distancia deslizada, o sea, cinco unidades a partir del nivel en que estaba antes. Nosotros *levantamos* el peso de un kilo solamente *tres* unidades y bajamos W kilos en *cinco* unidades. Por lo tanto, $W = \frac{3}{5}$ de kilo. Nótese que hemos deducido esto a partir de la *conservación de la energía* y no a partir de componentes de fuerzas. La habilidad es, sin embargo, relativa. Puede deducirse de una forma que es aún más brillante, descubierta por Stevinus e inscrita en su tumba. La figura 4-4 explica que debe ser $\frac{3}{5}$ de kilo, porque la cadena no da vuelta. Es evidente que la parte más baja de la cadena está equilibrada por sí misma, de modo que la tracción de los cinco pesos por un lado debe equilibrar la tracción de tres pesos por el otro, cualquiera que sea la proporción de los catetos. Ustedes ven, al observar este diagrama, que W debe ser $\frac{3}{5}$ de kilo. (Si logran un epitafio como éste en su lápida, van por buen camino.)

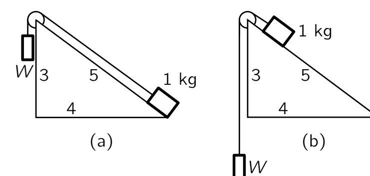


Figura 4-3. Plano inclinado.

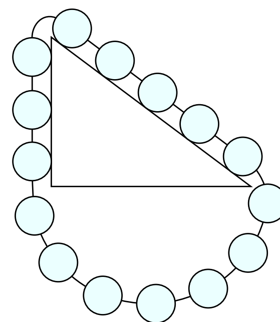


Figura 4-4. El epitafio de Stevinus.

Ilustremos ahora este principio de la energía con un problema más complicado, el tornillo que se muestra en la figura 4-5. Se usa una extensión de 20 centímetros de longitud para girar el tornillo que tiene 10 hilos por centímetro. Nos gustaría saber cuánta fuerza se necesita aplicar para levantar una tonelada. Si queremos levantar la tonelada un centímetro, digamos, entonces debemos girar la extensión 10 veces. Cuando da una vuelta recorre aproximadamente 126 centímetros. La extensión debe así recorrer 1 260 centímetros, y si usamos varias poleas, etc., estaríamos levantando nuestra tonelada con un peso menor W desconocido aplicado al extremo de la extensión. Así encontramos que W es de aproximadamente 1.6 kilos. Éste es un resultado de la conservación de la energía.

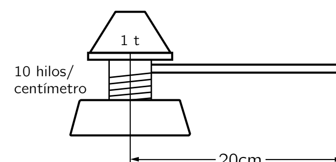


Figura 4-5. Un tornillo con extensión.

Consideremos ahora un ejemplo más complicado que se muestra en la figura 4-6. Una varilla o barra de 8 m de longitud que está soportada en un extremo. En la mitad de la barra hay un peso de 60 kg; a una distancia de 2 m del soporte hay un peso de 100 kg. ¿Cuánta fuerza debemos aplicar para levantar el extremo de la barra a fin de mantenerla en equilibrio, despreciando el peso de ésta? Supongamos que ponemos una polea en un extremo y colgamos un peso sobre la polea. ¿Cuál debe ser el valor del peso W para mantener el equilibrio? Imaginemos que el peso cae a una distancia arbitraria —para que sea más fácil para nosotros supongamos que baja 4 cm—; ¿a qué altura se elevarán las dos cargas? El centro se eleva 2 cm y el punto a un cuarto de la distancia del extremo fijo se eleva 1 cm. Por lo tanto, como el principio de que la suma de las alturas multiplicadas por los pesos no cambia, esto nos dice que el peso W por los 4 cm hacia abajo, más 60 kg por 2 cm hacia arriba, más 100 kg por 1 cm, tienen que sumar cero:

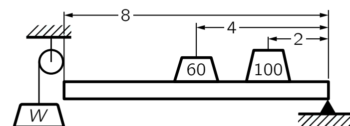


Figura 4-6. Barra cargada soportada en un extremo.

$$-4W + (2)(60) + (1)(100) = 0, \qquad W = 55 \text{ kg}. \tag{4.5}$$

Por lo tanto, debemos tener un peso de 55 kg para equilibrar la barra. De este modo podemos deducir las leyes del “equilibrio” —la estática de complicadas estructuras de puentes, etc. Este método se llama *principio de los trabajos virtuales*, porque para aplicar este argumento tuvimos que *imaginar* que la estructura se mueve un poco —aunque ni se mueva *realmente* ni se *pueda mover*—. Usamos este pequeño movimiento imaginario para aplicar el principio de conservación de la energía.

4-3 Energía cinética

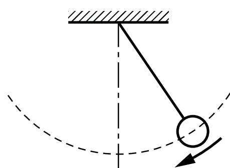


Fig. 4-7. Péndulo.

Para ilustrar otro tipo de energía consideremos un péndulo (figura 4-7). Si empujamos la masa hacia un lado y la soltamos, oscila de un lado hacia el otro. En su movimiento pierde altura cuando va desde ambos extremos hacia el centro. ¿A dónde se fue la energía potencial? La energía gravitacional desaparece cuando el cuerpo está abajo; sin embargo, subirá de nuevo. La energía gravitacional debe haberse convertido en otra forma. Evidentemente, es en virtud de su *movimiento* que es capaz de subir de nuevo, de modo que tenemos conversión de energía gravitacional en otra forma cuando el cuerpo llega al fondo.

Debemos obtener una fórmula para la energía de movimiento. Al recordar nuestros argumentos acerca de las máquinas reversibles podemos ver fácilmente que en el movimiento en la parte inferior debe haber una cantidad de energía que le permite subir a cierta altura, y que no tiene nada que ver con el *mecanismo* mediante el cual sube ni con la *trayectoria* según la cual sube. De modo que tenemos una fórmula de equivalencia parecida a la que escribimos para los bloques del niño. Tenemos otra forma de representar la energía. Es fácil decir cuál es. La energía cinética en el fondo es igual al peso multiplicado por la altura que puede alcanzar en correspondencia a su velocidad: $\mathrm{EC} = WH$. Lo que necesitamos es la fórmula que nos dé la altura mediante alguna regla que tenga que ver con el movimiento de los objetos. Si ponemos en marcha algo con cierta velocidad, por ejemplo verticalmente hacia arriba, alcanzará cierta altura; todavía no sabemos cuál es, pero depende de la velocidad —hay una fórmula para eso—. Entonces, para encontrar la fórmula de la energía cinética de un objeto que se mueve con velocidad V, debemos calcular la altura que alcanzaría y multiplicarla por el peso. Encontraremos luego que podemos escribirla de esta manera:

$$\mathrm{EC} = WV^2/2g. \tag{4.6}$$

Por supuesto, el hecho de que el movimiento tenga energía no tiene nada que ver con el de que estemos en un campo gravitacional. No importa *de dónde* vino el movimiento. Ésta es una fórmula general, para diversas velocidades. Las fórmulas (4.3) y (4.6) son ambas aproximadas, la primera porque es incorrecta cuando las alturas son grandes, es decir, cuando las alturas son tan grandes que la gravedad se debilita; la segunda debido a la corrección relativista para grandes velocidades. Sin embargo, cuando finalmente obtengamos la fórmula exacta para la energía, entonces la ley de conservación de la energía es correcta.

4-4 Otras formas de energía

Podemos, de este modo, continuar ilustrando la existencia de energía bajo otras formas. Primero, consideremos la energía elástica. Si comprimimos un resorte hacia abajo, debemos hacer cierto trabajo porque, una vez hecho eso, podemos levantar pesos con él. Por lo tanto, en su condición comprimida tiene la posibilidad de hacer cierto trabajo. Si calculáramos las sumas de los productos por las alturas, la ley no se verificaría —debemos agregar algo más para tomar en cuenta el hecho de que el resorte está bajo tensión—. Energía elástica es la receta para un resorte cuando está comprimido. ¿Cuánta energía es? Si soltamos la energía elástica, a medida que el resorte pasa por el punto de equilibrio, se convierte en energía cinética y ésta pasa alternativamente por compresiones o estiramientos del resorte y energía cinética de movimiento. (Hay, además, cierta energía gravitacional que entra y sale, pero podemos hacer este experimento "de costado" si lo deseamos). Se mantendrá en movimiento hasta que frene por pérdidas. ¡Ah! Hemos estado trampeando todo el tiempo, poniendo pequeños pesos para mover cosas, o diciendo que las máquinas son reversibles, o que se mueven permanentemente; pero podemos ver que las cosas se detienen a la larga. ¿Dónde está la energía cuando el resorte ha terminado de moverse de arriba a abajo? Esto introduce *otra* forma de energía: la *energía térmica*.

Dentro de un resorte o de una palanca hay cristales formados por muchos átomos y con gran cuidado y delicadeza en la disposición de las partes uno

puede tratar de ajustar las cosas de modo que ruede sobre algo sin que ninguno de los átomos verifique agitación alguna. Pero uno debe tener mucho cuidado. Ordinariamente cuando las cosas ruedan, hay sacudimiento y agitación debido a las irregularidades del material y los átomos comienzan a menearse en el interior. Así perdemos la pista de esa energía; encontramos que los átomos están meneándose en el interior de una manera al azar y confusa después que el movimiento se detuvo. Aún hay energía cinética, por cierto, pero no está asociada a un movimiento visible. ¡Estamos soñando! ¿Cómo *sabemos* que aún hay energía cinética? Resulta que con termómetros pueden encontrar que, de hecho, el resorte o la palanca están *más calientes* y que hay realmente un incremento de la energía cinética en una cantidad definida. Llamamos *energía térmica* a esta energía, pero sabemos que ésta no es realmente una nueva forma, es justamente energía cinética —movimiento interno—. (Una de las dificultades con todos estos experimentos que hacemos con materia a gran escala es que no podemos demostrar realmente la conservación de la energía y no podemos construir realmente nuestras máquinas reversibles, porque en cada momento movemos una gran masa de sustancia y los átomos no permanecen absolutamente inalterados, y así una cierta cantidad de movimiento al azar ocurre dentro del sistema atómico. No podemos verlo, pero podemos medirlo con un termómetro, etcétera).

Hay muchas otras formas de energía y, por supuesto, no podemos describirlas con más detalle ahora. Existe la energía eléctrica, que tiene que ver con el empuje y arrastre de cargas eléctricas. Existe la energía radiante, la energía de la luz, que sabemos es una forma de la energía eléctrica, porque la luz puede representarse como oscilaciones en el campo electromagnético. Existe la energía química, la energía que es liberada en las reacciones químicas. Realmente, la energía elástica es, hasta cierto punto, como la energía química, porque la energía química es la energía de atracción de los átomos, de uno al otro, y así es la energía elástica. Nuestro conocimiento moderno es el siguiente: la energía química consta de dos partes, energía cinética de los electrones en el interior de los átomos, así parte de ella es cinética, y energía eléctrica de interacción de los electrones y protones —por lo tanto, el resto de ella es eléctrico—. Enseguida llegamos a la energía nuclear, la energía involucrada en el arreglo de las partículas dentro del núcleo, y tenemos fórmulas para eso, pero no tenemos sus leyes fundamentales. Sabemos que no es eléctrica, no es gravitacional y no es puramente cinética, pero no sabemos lo que es. Parece ser una forma adicional de energía. Finalmente, asociada con la teoría de la relatividad, hay una modificación de las leyes de la energía cinética, o como quieran llamarla, de modo tal que la energía cinética está combinada con otra cosa llamada *energía de masa*. Un objeto tiene energía a partir de su sola *existencia*. Si yo tengo un positrón y un electrón que permanecen quietos sin hacer nada —sin importar la gravedad, sin importar cualquier otra cosa— y se juntan y desaparecen, se libera energía radiante en una cantidad definida, y la cantidad puede calcularse. Todo lo que necesitamos saber es la masa del objeto. No depende de lo que sea —podemos hacer desaparecer dos cosas y obtener cierta cantidad de energía—. La fórmula fue encontrada primero por Albert Einstein; ésta es: $E = mc^2$.

Es evidente a partir de nuestra discusión que la ley de conservación de la energía es enormemente útil para hacer análisis, como lo hemos ilustrado con unos pocos ejemplos sin conocer todas las fórmulas. Si tuviéramos todas las fórmulas para todas las formas de energía, podríamos analizar cuántos procesos deberían verificarse sin tener que entrar en detalles. Por eso las leyes de conservación son muy interesantes. La cuestión que naturalmente surge es qué otras leyes de conservación hay en física. Hay otras dos leyes de conservación que son análogas a la conservación de la energía. Una se llama la conservación del momentum lineal. La otra se llama la conservación del momentum angular. Nosotros averiguaremos más acerca de éstas más adelante. En último análisis no entendemos en profundidad las leyes de conservación. No entendemos la ley de conservación de la energía. No entendemos la energía como cierto número de pequeñas gotas. Puede que hayan oído que los fotones se manifiestan como gotas y que la energía de un fotón es la constante de Planck multiplicada por

la frecuencia. Esto es cierto, pero ya que la frecuencia de la luz puede tomar cualquier valor, no hay ninguna ley que diga que la energía tenga que ser una cierta cantidad definida. A diferencia de los bloques de Daniel, puede haber cualquier cantidad de energía, por lo menos como se entiende actualmente. De manera que, por ahora, no entendemos esta energía como contar algo, sino como una cantidad matemática, lo que es una circunstancia abstracta y más bien peculiar. En mecánica cuántica resulta que la conservación de la energía está muy estrechamente relacionada con otra importante propiedad del mundo, *las cosas no dependen del tiempo absoluto*. Podemos montar un experimento en un momento dado y probarlo, y luego hacer el mismo experimento en un momento posterior, y él se desarrolla exactamente en la misma forma. Si esto es estrictamente cierto o no, no lo sabemos. Si suponemos que *es* cierto y agregamos los principios de la mecánica cuántica, entonces podemos deducir el principio de la conservación de la energía. Es una cosa más bien sutil e interesante y no es fácil de explicar. Las otras leyes de conservación están también ligadas entre sí. La conservación del momentum está asociada en mecánica cuántica con la proposición que no importa *dónde* se haga el experimento; los resultados serán siempre los mismos. Así como la independencia en el espacio tiene que ver con la conservación del momentum, la independencia en el tiempo tiene que ver con la conservación de la energía, y finalmente si *giráramos* nuestros aparatos, esto tampoco implica diferencia, de modo que la invariancia del mundo a la orientación angular está relacionada con la conservación del *momentum angular*. Además de éstas, hay tres leyes de conservación que son exactas hasta donde lo podemos afirmar hoy día, que *son* mucho más simples de entender, porque son del mismo tipo que contar bloques.

La primera de las tres es la *conservación de la carga*, y que sencillamente significa que ustedes cuentan las cargas positivas, menos las cargas negativas que tienen, y el número no cambia nunca. Pueden deshacerse de una positiva con una negativa, pero no crean ningún exceso neto de positivas sobre negativas. Hay otras dos leyes análogas a ésta —una es la llamada *conservación de bariones*—. Hay cierto número de partículas extrañas (un neutrón y un protón, por ejemplo) que se llaman bariones. En cualquier reacción en la naturaleza, si contamos cuántos bariones intervienen en un proceso, el número de bariones* que resulta será exactamente el mismo. Hay otra ley, la *conservación de leptones*. Podemos decir que ese grupo de partículas llamadas leptones es: el electrón, el muon y el neutrino. Existe un antielectrón, que es un positrón, esto es, -1 leptón. Al contar el número total de leptones en una reacción resulta que el número que entra y sale nunca cambia, al menos por lo que sabemos hasta el presente.

Éstas son las seis leyes de conservación, tres de ellas sutiles, involucrando espacio y tiempo, y tres de ellas simples, en el sentido de contar cosas.

Con relación a la conservación de la energía debiéramos notar que la energía *disponible* es otro asunto —hay mucha agitación de los átomos del agua de mar, puesto que el mar tiene cierta temperatura, pero es imposible reunirlos en un movimiento definido sin tomar energía de cualquier otra parte—. Es decir, aunque sabemos, de hecho, que la energía se conserva, la energía disponible para la utilidad humana no se conserva tan fácilmente. Las leyes que gobiernan la cantidad de energía disponible se llaman *leyes de la termodinámica* y encierran un concepto llamado entropía para los procesos termodinámicos irreversibles.

Finalmente, reparamos en el problema de dónde podemos obtener nuestras fuentes de energía hoy día. Nuestros abastecimientos de energía provienen del Sol, la lluvia, el carbón, el uranio y el hidrógeno. El Sol forma la lluvia y también el carbón, de modo que todo esto proviene del Sol. Aunque la energía se conserva, la naturaleza no parece interesada en ello; libera gran cantidad de energía desde el Sol, pero sólo una parte en 2 000 millones cae sobre la Tierra. La naturaleza conserva la energía, pero, realmente, no le importa; derrocha mucha en todas direcciones. Ya hemos obtenido energía del uranio, podemos obtenerla también del hidrógeno, pero actualmente sólo en una condición explosiva y peligrosa.

* Contando los antibariones como -1 barión.

Si pudiera ser controlada en reacciones termonucleares, resulta que la energía que pueda obtenerse a partir de 10 litros de agua por segundo es igual a toda la potencia eléctrica generada en los Estados Unidos. ¡Con 568 litros de agua corriente por minuto tienen suficiente combustible para abastecer toda la energía que se usa hoy día en los Estados Unidos! Por lo tanto, concierne a los físicos resolver cómo liberarnos de la necesidad de tener energía. Esto puede hacerse.

5

Tiempo y distancia

5-1 El movimiento

En este capítulo consideraremos algunos aspectos de los conceptos de *tiempo* y *distancia.* Se ha puesto énfasis desde un comienzo que la física, como todas las ciencias, depende de la *observación.* Se puede decir, también, que el desarrollo de las ciencias físicas en su forma actual ha dependido en gran medida del énfasis que se ha puesto en hacer observaciones *cuantitativas.* Sólo mediante observaciones cuantitativas puede llegar uno a relaciones cuantitativas, que son el corazón de la física.

A muchas personas les gusta colocar el comienzo de la física en el trabajo hecho hace 350 años por Galileo y llamarlo el primer físico. Hasta esa época el estudio del movimiento había sido filosófico y basado en argumentos que podrían idearse con la cabeza. La mayoría de los razonamientos habían sido presentados por Aristótles y otros filósofos griegos y se tomaban como "probados". Galileo era escéptico, e hizo un experimento sobre el movimiento, que fue esencialmente el siguiente: hizo rodar una bola hacia abajo en un plano inclinado y observó su movimiento. Sin embargo, no sólo observó; midió *la distancia* que recorrió la bola y en *cuánto tiempo.*

El modo de medir una distancia era bien conocido mucho antes de Galileo, pero no había modos precisos de medir el tiempo, particularmente tiempos cortos. Aunque después diseñó relojes más satisfactorios (aunque no como los que conocemos ahora), el primer experimento de Galileo sobre el movimiento fue hecho usando su pulso para contar intervalos iguales de tiempo. Hagamos lo mismo.

Podemos contar los latidos del pulso mientras la bola rueda hacia abajo sobre la pista: "uno... dos ... tres... cuatro... cinco... seis... siete... ocho... " Pedimos a un amigo que haga una pequeña marca en la ubicación de la bola para cada cuenta; podemos entonces medir la *distancia* que se movió la bola desde el punto en que se soltó en uno, o dos, o tres, etc., intervalos iguales de tiempo. Galileo expresó el resultado de *sus* observaciones de este modo: si la ubicación de la bola es marcada a 1, 2, 3, 4... unidades de tiempo desde el instante en que se suelta, esas marcas distan del punto de partida en proporción a los números 1, 4, 9, 16... Hoy diríamos que la distancia es proporcional al cuadrado del tiempo.

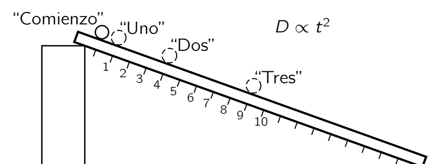


Figura 5-1. Una bola baja rodando por una pista inclinada.

$$D \propto t^2.$$

El estudio del *movimiento*, que es básico para todos los físicos, implica las preguntas ¿dónde? y ¿cuándo?

5-2 El tiempo

Consideremos primero lo que queremos decir por *tiempo.* ¿Qué *es* el tiempo? Sería estupendo encontrar una buena definición del tiempo. El diccionario Webster define "un tiempo" como "un periodo", y este último como "un tiempo", lo que no parece ser muy útil. Quizá debiéramos decir: "tiempo es lo que ocurre, cuando nada más ocurre". Lo que tampoco nos lleva muy lejos. Puede ser igualmente bueno que enfrentemos el hecho de que el tiempo es, probablemente, una de las cosas que no podemos definir (en el sentido del diccionario), y sólo decir que es lo que ya sabemos que es: ¡es cuanto esperamos!

De todos modos, lo que realmente importa no es *definir* el tiempo, sino cómo medirlo. Una manera de medir el tiempo es utilizar algo que ocurra una y otra

vez de modo regular —algo que sea *periódico*—. Por ejemplo un día. Un día parece ocurrir una y otra vez. Pero cuando comienzan a pensar acerca de ello, bien pueden preguntar: "¿Son los días periódicos? ¿Son regulares? ¿Son todos los días de la misma duración?" Ciertamente, uno tiene la impresión que los días de verano son más largos que los de invierno. Por supuesto, algunos días de invierno parecen ser desagradablemente largos si uno está muy aburrido. Seguramente han oído decir a alguien: "¡Caramba, éste sí que ha sido un día largo!"

Sin embargo, parece ser que los días son casi del mismo largo *en promedio*. ¿Hay algún modo para verificar si los días son del mismo largo —sea de un día al siguiente— o por lo menos en promedio? Es posible comparar con algún otro fenómeno periódico. Veamos cómo puede hacerse tal comparación usando un reloj de arena. Con un reloj de arena podemos "crear" un suceso periódico siempre que tengamos a alguien ocupado durante el día y la noche para girarlo cuando el último grano de arena haya caído.

Entonces podríamos contar las vueltas del reloj de arena desde cada mañana a la siguiente. Encontraríamos esta vez que el número de "horas" (es decir, el número de vueltas que damos al reloj de arena) no fue el mismo cada "día". Podríamos desconfiar del sol, o del reloj de arena o de ambos. Después de pensar un poco, puede ocurrírsenos contar las "horas" de mediodía a mediodía (el mediodía *no* está definido aquí como las 12:00 horas, sino como el instante en que el sol está en su punto más alto). Encontraríamos esta vez que el número de horas de cada día es el mismo.

Tenemos ahora alguna confianza en que tanto la "hora" como el "día" tienen una periodicidad regular, es decir, indican sucesivos intervalos iguales de tiempo, aunque no hemos *probado* que cada uno sea "realmente" periódico. Alguien puede preguntar si no pudiera haber algún ser omnipotente que hiciera disminuir el flujo de arena cada noche y aumentarlo durante el día. Nuestro experimento, por supuesto, no nos da una respuesta sobre esta clase de preguntas. Todo lo que podemos decir es que encontramos que una regularidad en una cosa lleva consigo una regularidad en la otra. Podemos sólo decir que basamos nuestra *definición* del tiempo en la repetición de algún evento aparentemente periódico.

5-3 Tiempos cortos

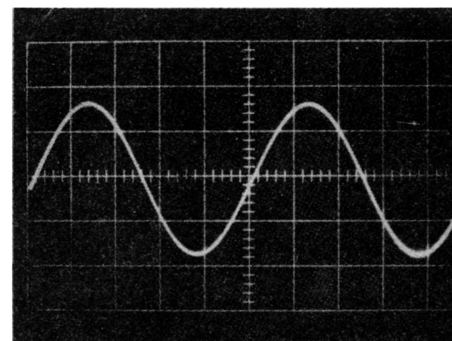
Podríamos indicar ahora que en el proceso de verificar la reproductibilidad del día, hemos recibido un importante subproducto. Hemos encontrado la manera de medir en forma más precisa *fracciones* de día. Hemos encontrado un modo de contar el tiempo en pequeñas partes. ¿Podemos llevar más lejos este proceso y aprender a medir intervalos de tiempo aún más pequeños?

Galileo determinó que un péndulo siempre oscila en intervalos iguales de tiempo, siempre que el tamaño de la oscilación se mantenga pequeño. Un experimento que compare el número de oscilaciones de un péndulo en una "hora" demuestra que éste es efectivamente el caso. En esta forma podemos marcar fracciones de una hora. Si usamos un dispositivo mecánico para contar las oscilaciones —y para mantenerlas—, tenemos el reloj de péndulo de nuestros abuelos.

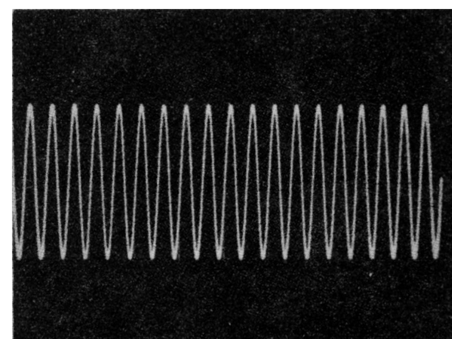
Convengamos que si nuestro péndulo oscila 3 600 veces en una hora (y si hay 24 de tales horas en un día) llamaremos a cada periodo del péndulo un "segundo". Hemos dividido entonces nuestra unidad original de tiempo en aproximadamente 10^5 partes. Podemos aplicar los mismos principios para dividir el segundo en intervalos más y más pequeños. Se darán cuenta de que no es práctico hacer péndulos mecánicos que funcionen arbitrariamente rápido, pero ahora podemos hacer péndulos *eléctricos* llamados osciladores, que pueden proporcionar sucesos periódicos con un tiempo de oscilación muy pequeño. En estos osciladores electrónicos hay una corriente eléctrica que oscila en forma análoga a la oscilación de la lenteja del péndulo.

Podemos hacer una serie de tales osciladores electrónicos cada uno con un periodo 10 veces más corto que el precedente. Podemos "calibrar" cada oscilador con el siguiente más lento, contando el número de oscilaciones que hace por una

oscilación del más lento. Cuando el periodo de oscilación de nuestro reloj es menor que una fracción de segundo, no podemos contar las oscilaciones sin la ayuda de algún dispositivo que extienda nuestro poder de observación. Un dispositivo tal es un osciloscopio de haz electrónico, el cual actúa como una especie de microscopio para tiempos cortos. Este dispositivo traza en una pantalla fluorescente un gráfico de la corriente eléctrica (o el voltaje) en función del tiempo. Al conectar el osciloscopio a dos de nuestros osciladores uno después del otro, de modo que primero hace un gráfico de la corriente en uno de nuestros osciladores y después de la corriente en el otro, obtenemos dos gráficos como los que se muestran en la figura 5-2. Podemos determinar fácilmente el número de periodos del oscilador más rápido en un periodo del oscilador más lento.



(a)



(b)

Figura 5-2. Dos vistas de la pantalla de un osciloscopio. En (a) el osciloscopio está conectado a un oscilador; en (b) está conectado a un oscilador con un periodo de un décimo de duración.

Con las modernas técnicas electrónicas se han construido osciladores que tienen periodos tan pequeños como de 10^{-12} segundos, y han sido calibrados (por métodos de comparación como los que hemos descrito) en términos de nuestra unidad patrón de tiempo, el segundo. Con la invención y perfeccionamiento del "láser" o amplificador de luz, en estos últimos años, ha llegado a ser posible construir osciladores con periodos aún más cortos que 10^{-12} segundos; pero no ha sido todavía posible calibrarlos mediante los métodos que hemos descrito, aunque no hay duda de que pronto será posible.

Se han medido intervalos de tiempo más pequeños que 10^{-12} segundos, pero mediante una técnica diferente. En efecto, se ha usado una *definición* diferente del "tiempo". Una manera ha sido observar la *distancia* entre dos sucesos para un objeto en movimiento. Si, por ejemplo, las luces delanteras de un automóvil en movimiento se prenden y después se apagan, podemos calcular *cuánto tiempo* estuvieron prendidas las luces, si sabemos *dónde* fueron encendidas y apagadas y la velocidad con que se movió el automóvil. El tiempo es la distancia durante la cual las luces estuvieron encendidas, dividida entre la velocidad.

En estos últimos años justamente se utilizó una técnica así para medir el tiempo de vida del mesón π^0. Al observar en un microscopio las pequeñas trazas dejadas en una emulsión fotográfica, en la cual los mesones π^0 habían sido creados, se vio que un mesón π^0 (se sabe que viaja a una velocidad cercana a la de la luz) se movió una distancia de 10^{-7} metros en promedio antes de desintegrarse. Vivió sólo durante unos 10^{-16} segundos. Deberá ponerse énfasis que aquí hemos utilizado una definición un poco diferente del "tiempo" que las anteriores. Mientras no haya contradicciones en nuestra comprensión, nos sentimos bastante seguros de que nuestras definiciones son suficientemente equivalentes.

Extendiendo nuestras técnicas aún más —y si es necesario, nuestras definiciones— podemos inferir el tiempo de duración de acontecimientos aún más rápidos. Podemos hablar del periodo de una vibración nuclear. Podemos hablar del tiempo de vida de las recientemente descubiertas resonancias extrañas (partículas) mencionadas en el capítulo 2. Su vida entera dura un lapso de sólo 10^{-24} segundos, el tiempo que aproximadamente tardaría la luz (la que se mueve a la mayor de las velocidades conocidas) en atravesar el núcleo de hidrógeno (el más pequeño de los objetos conocidos).

¿Qué hay de tiempos aún más pequeños? ¿Existe el "tiempo" en una escala aún más pequeña? ¿Tiene sentido hablar de tiempos menores si no los podemos medir —o quizás incluso ni pensar en ello sensatamente—, algo que sucede en un tiempo aún más pequeño? Quizá no. Éstos son algunos de los problemas no resueltos que ustedes estarán preguntando y quizá contestando en los próximos 20 o 30 años.

5-4 Tiempos largos

Consideremos ahora tiempos mayores que un día. La medida de tiempos largos es fácil; sólo contamos los días siempre que haya alguien que lleve la cuenta. Desde luego encontramos que hay otra periodicidad natural: el año, de alrededor de 365 días. También hemos descubierto que la naturaleza entrega algunas veces un contador para los años, en forma de anillos de los árboles o sedimentos del fondo de los ríos. En algunos casos podemos emplear esos

Tiempos

Años	Segundos		Vida media de
		? ? ? ? ? ? ? ?	
	10^{18}	Edad del universo	
		Edad de la Tierra	U^{238}
10^{9}			
	10^{15}		
		Primeros hombres	
10^{6}			
	10^{12}	Edad de las pirámides	
			Ra^{226}
10^{3}			
		Edad de los Estados Unidos	
	10^{9}	Vida de un hombre	H^{3}
1			
	10^{6}		
		Un día	
	10^{3}	La luz va del Sol a la Tierra	Neutrón
	1	Un latido del corazón	
	10^{-3}	Periodo de una onda sonora	
	10^{-6}	Periodo de una onda de radio	Muon
			Mesón $\pi^{\pm}$
	10^{-9}	La luz viaja 30 centímetros	
	10^{-12}	Periodo de la rotación molecular	
	10^{-15}	Periodo de la vibración atómica	
			Mesón π^{0}
	10^{-18}	La luz atraviesa un átomo	
	10^{-21}		
		Periodo de la vibración nuclear	
	10^{-24}	La luz atraviesa un núcleo	Partícula extraña
		? ? ? ? ? ? ? ?	

marcadores naturales del tiempo para determinar el lapso que ha transcurrido desde algún acontecimiento primitivo.

Cuando no podemos contar los años para la medida de tiempos largos, debemos buscar otras formas de medir. Una de las que ha tenido más éxito es usar como "reloj" un material radiactivo. En este caso no tenemos un acontecer periódico como el día o el péndulo, sino un nuevo tipo de "regularidad". Encontramos que la radiactividad de una muestra de material particular decrece la misma *fracción* para sucesivos aumentos iguales en su edad. Si trazamos un gráfico de la radiactividad observada como función del tiempo (digamos en días), obtenemos una curva como la que se muestra en la figura 5-3. Observamos que si la radiactividad decrece a la mitad en T días (llamada "vida media"), entonces decrece a un cuarto en otros T días, y así sucesivamente. En un intervalo de tiempo arbitrario t hay t/T "vidas medias", entonces la fracción que queda después de este tiempo t es $(\frac{1}{2})^{t/T}$.

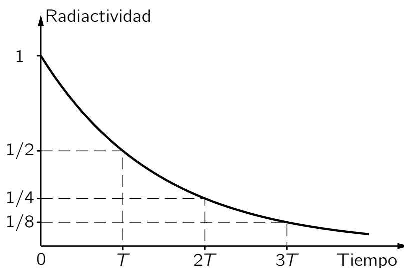


Figura 5-3. El decrecimiento de la radiactividad en el tiempo. La actividad decrece a la mitad en cada "vida media", T.

Si supiéramos que un pedazo de material, digamos un pedazo de madera, contenía una cantidad A de material radiactivo cuando se formó, y encontráramos por una medida directa que ahora contiene una cantidad B, podríamos calcular la edad del objeto t, resolviendo la ecuación

$$(\tfrac{1}{2})^{t/T} = B/A.$$

Afortunadamente hay casos en los cuales podemos conocer la cantidad de radiactividad que tenía un objeto cuando se formó. Sabemos, por ejemplo, que el dióxido de carbono en el aire contiene una cierta pequeña fracción del isótopo radiactivo del carbono, C^{14} (reabastecido continuamente por la acción de los rayos cósmicos). Si medimos el contenido *total* de carbono de un objeto, sabemos que cierta reacción de esa cantidad era originalmente el C^{14} radiactivo; conocemos, por lo tanto, la cantidad inicial A a usar en la fórmula anterior. El C^{14} tiene una vida media de 5 000 años. Con medidas cuidadosas podemos medir la cantidad

que queda después de 20 vidas medias o algo así, y por lo tanto “fechar” objetos orgánicos que crecieron hace 100 000 años.

Nos gustaría conocer, y creemos conocer, la vida de cosas aún más antiguas. Mucho de nuestro conocimiento está basado en las medidas de otros isótopos radioactivos que tienen diferentes vidas medias. Si hacemos medidas con un isótopo de vida media mayor, entonces podremos medir tiempos mayores. Por ejemplo, el uranio tiene un isótopo cuya vida media es alrededor de 10^9 años, de modo que si algún material se formó con uranio en él hace 10^9 años, hoy en día sólo quedaría la mitad del uranio. Cuando el uranio se desintegra, se convierte en plomo. Consideren un pedazo de roca que se formó mucho tiempo atrás en algún proceso químico. El plomo, siendo de naturaleza química diferente al uranio, aparecería en una parte de la roca y el uranio aparecería en otra parte de la roca. El uranio y el plomo estarían separados. Si observamos hoy ese pedazo de roca, donde debiera haber sólo uranio encontraremos ahora una cierta fracción de uranio y una cierta fracción de plomo. Por comparación de estas fracciones podemos decir cuál es el porcentaje de uranio que desapareció y se transformó en plomo. Con este método se ha determinado la edad de ciertas rocas, resultando ser de varios miles de millones de años. Una extensión de este método que no usa rocas particulares, pero que observa el uranio y plomo en los océanos y usa promedios sobre la Tierra, ha sido utilizada para diagnosticar (en los últimos años) que la edad de la Tierra misma es aproximadamente de 4 500 millones de años.

Es alentador que se haya encontrado que la edad de la Tierra sea la misma que la edad de los meteoritos que caen sobre ella, según se acuerda por el método del uranio. Parece que la Tierra surgió de las rocas que flotan en el espacio, y que los meteoritos, muy probablemente, son algo de aquel material que quedó. En un cierto instante, hace más de 5 000 millones de años, comenzó el universo. Se cree ahora que por lo menos nuestra parte del universo tuvo su comienzo hace cerca de 10 000 a 20 000 millones de años. No sabemos lo que ocurrió antes de esto. De hecho, podemos muy bien preguntar de nuevo: ¿tiene algún sentido la pregunta? ¿Tienen los tiempos anteriores algún significado?

5-5 Unidades y patrones de tiempo

Hemos dado a entender que es conveniente comenzar con alguna unidad patrón de tiempo, digamos un día o un segundo, y referir todos los otros tiempos a algún múltiplo o fracción de esta unidad. ¿Qué tomaremos como nuestro patrón básico de tiempo? ¿Tomaremos el pulso humano? Si comparamos los pulsos, encontramos que ellos varían mucho. Al comparar dos relojes se encuentra que no varían tanto. Ustedes podrían entonces decir: bien, tomemos un reloj. ¿Pero el reloj de quién? Hay una historia de un niño suizo, quería que todos los relojes de la ciudad sonaran al mediodía al mismo tiempo. Entonces fue a todas partes tratando de convencer a todos del valor de esto. ¡Todos pensaron que era una idea maravillosa que todos los otros relojes sonaran al mediodía cuando lo hiciera el propio! Es más bien difícil decidir el reloj de quién debiéramos tomar como patrón. Afortunadamente todos compartimos un reloj —la Tierra—. Desde hace mucho tiempo se ha tomado el periodo rotacional de la Tierra como el patrón básico del tiempo. Sin embargo, a medida que se han ido haciendo medidas más y más precisas se ha encontrado que la rotación de la Tierra no es exactamente periódica cuando se mide con los mejores relojes. Estos “mejores” relojes son aquellos para los cuales tenemos razón de creer que son precisos, porque coinciden entre sí. Creemos ahora que, por varias razones, algunos días son más largos que otros, algunos días son más cortos y que en promedio el periodo de la Tierra llega a ser un poco más largo a medida que los siglos pasan.

Hasta hace muy poco no hemos encontrado nada mejor que el periodo de la Tierra, así todos los relojes se han referido a la longitud del día y se ha definido el segundo como 1/86 400 de un día promedio. Recientemente hemos estado ganando experiencia con algunos osciladores naturales que creemos ahora podrían proporcionar una referencia de tiempo más constante que la Tierra y que también

están basados en un fenómeno natural disponible para todos. Éstos son los así llamados “relojes atómicos”. Su periodo interno básico es el de una vibración atómica que es muy insensible a la temperatura o cualesquiera otros efectos externos. Estos relojes se mantienen en hora con una precisión de una parte en 10^9, o mejor. Dentro de los dos últimos años un relojes atómico perfeccionado que opera basado en la vibración del átomo de hidrógeno ha sido diseñado y construido por el profesor Norman Ramsey de la Universidad de Harvard. Él piensa que este reloj podría ser 100 veces más preciso aún. Las mediciones que se están realizando mostrarán si esto es verdad o no.

Podemos esperar que ya que ha sido posible construir relojes mucho más precisos que el tiempo astronómico, habrá pronto un acuerdo entre los científicos para definir la unidad de tiempo en términos de un reloj atómico estándar.

5-6 Distancias grandes

Volvamos ahora al problema de la *distancia*. ¿A qué distancia están o qué tamaño tienen las cosas? Cualquiera sabe que la manera de medir distancias es empezar con una regla y contar. O empezar con un pulgar y contar. Empiezan con una unidad y cuentan. ¿Cómo mide uno cosas más pequeñas? ¿Cómo podemos subdividir una distancia? De la misma manera que subdividimos el tiempo: tomamos una unidad menor y contamos el número de tales unidades que se necesitan para obtener la unidad mayor. Así podemos medir longitudes más y más pequeñas.

Pero no siempre entendemos por distancia lo que se obtiene contando sucesivamente con una varilla métrica. Sería difícil medir la distancia horizontal entre las cumbres de dos montañas usando sólo una varilla métrica. Hemos encontrado por experiencia que la distancia puede medirse de otro modo: por triangulación. Aunque esto significa que realmente estamos usando una definición diferente de distancia, concuerdan cuando ambas pueden usarse. El espacio es más o menos lo que Euclides pensó que era, de modo que los dos tipos de definición de distancia concuerdan. El hecho que concuerden en la Tierra nos da confianza para usar la triangulación para distancias aún más grandes. Por ejemplo, pudimos usar ésta estructura para medir la altura del primer Sputnik. Encontramos que estaba aproximadamente a 5×10^5 metros de altura. Con mediciones más cuidadosas puede medirse de la misma manera la distancia a la Luna. Dos telescopios en diferentes lugares sobre la Tierra nos pueden dar los dos ángulos que necesitamos. De este modo se ha encontrado que la Luna está a 4×10^8 metros de distancia.

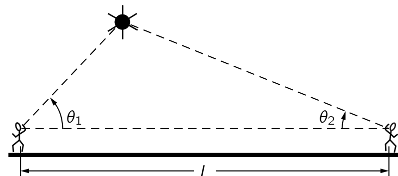


Figura 5-4. La altura de un Sputnik se determina por triangulación.

No podemos hacer lo mismo con el Sol, o al menos nadie ha sido capaz todavía. La exactitud con que se puede enfocar cierto punto del Sol y con que se puede medir ángulos no es lo suficientemente buena para permitirnos medir la distancia al Sol. ¿Cómo podemos medir entonces la distancia al Sol? Debemos generalizar la idea de triangulación. Medimos las distancias relativas de todos los planetas por observaciones astronómicas de las posiciones aparentes de los planetas y obtenemos un cuadro del sistema solar con las distancias propias *relativas* de cada cosa, pero no con distancias *absolutas*. Se requiere, entonces, una medida absoluta que ha sido obtenida de varios modos. Una de esas formas, que se creía hasta hace poco ser la más exacta, consistió en medir la distancia desde la Tierra a Eros, uno de los pequeños planetoides que pasa cerca de la Tierra de vez en cuando. Por triangulación sobre este pequeño objeto, se pudo obtener la escala de medida requerida. Conociendo las distancias relativas del resto, podemos, por ejemplo, indicar la distancia de la Tierra al Sol o de la Tierra a Plutón.

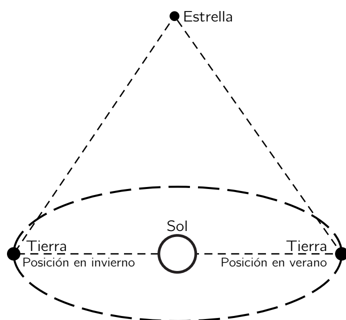


Figura 5-5. La distancia de estrellas cercanas se puede medir por triangulación, usando el diámetro de la trayectoria de la Tierra como línea de base.

En los últimos años ha habido un gran progreso en nuestro conocimiento de la escala del sistema solar. En el Jet Propulsión Laboratory se midió muy exactamente la distancia de la Tierra a Venus mediante una observación directa con radar. Esto, por supuesto, es un tipo todavía diferente de distancia deducida. Decimos que conocemos la velocidad a que viaja la luz (y, por lo tanto, a la que viajan las ondas de radar), y suponemos que es la misma velocidad en todas partes entre la Tierra y Venus. Enviamos la onda de radio, y contamos el tiempo hasta que la onda reflejada regrese. A partir del *tiempo* inferimos una *distancia*,

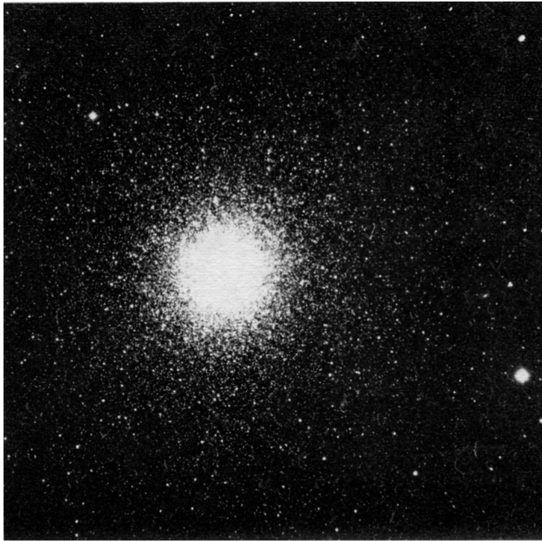


Figura 5-6. Un cúmulo de estrellas cerca del centro de nuestra galaxia. Su distancia a la tierra es 30 000 años luz, o alrededor de 3×10^{20} metros.

asumiendo que conocemos la velocidad. Tenemos en realidad otra definición de una medida de distancia.

¿Cómo medimos la distancia a una estrella que está mucho más lejos? Afortunadamente, podemos volver a nuestro método de triangulación, ya que la Tierra en su movimiento alrededor del Sol nos da una gran línea de base para medir los objetos que están fuera del sistema solar. Si enfocamos un telescopio hacia una estrella en verano y en invierno, esperaríamos determinar con precisión suficiente estos dos ángulos para poder medir la distancia a una estrella.

¿Qué pasa si las estrellas están demasiado lejos como para usar la triangulación? Los astrónomos están siempre inventando nuevos modos de medir distancias. Ellos encuentran, por ejemplo, que pueden estimar el tamaño y brillo de una estrella por su color. Se ha medido el color y brillo de muchas estrellas cercanas —cuyas distancias se conocen por triangulación—, y se ha encontrado que hay una pequeña relación entre el color y el brillo propio de las estrellas (en la mayoría de los casos). Si uno posee medidas de color de una estrella distante, se puede usar la relación color-brillo para determinar el brillo propio de la estrella. Midiendo lo brillante que nos *parece* la estrella desde la Tierra (o tal vez deberíamos decir cuán *débil* parece), podemos calcular lo alejada que está. (Para cierto brillo propio, el brillo aparente decrece con el cuadrado de la distancia). Una hermosa confirmación de lo correcto de este método para medir distancias estelares está dada por los resultados obtenidos para grupos de estrellas, conocidos como cúmulos globulares. Una fotografía de un grupo tal se muestra en la figura 5-6. Sólo con mirar la fotografía uno se convence de que estas estrellas están todas juntas. El mismo resultado se obtiene a partir de las medidas de distancia por el método color-brillo.

Un estudio de muchos cúmulos globulares da importante información. Se encuentra que hay una alta concentración de tales cúmulos en cierta parte del cielo, y que la mayoría de ellos está aproximadamente a la misma distancia de nosotros. Juntando esta información con otra evidencia, concluimos que esta concentración de cúmulos marca el centro de nuestra galaxia. Conocemos entonces la distancia al centro de la galaxia —alrededor de 10^{20} metros—.

Conociendo el tamaño de nuestra propia galaxia, tenemos una clave para la medida de distancias aún mayores —la distancia a otras galaxias—. La figura 5-7 es una fotografía de una galaxia que tiene una forma muy parecida a la nuestra. Probablemente es del mismo tamaño también. (Otras evidencias apoyan la idea que todas las galaxias son más o menos del mismo tamaño). Si es del mismo tamaño que la nuestra, podemos indicar su distancia. Medimos el ángulo



Figura 5-7. Una galaxia en espiral, como la nuestra. Suponiendo que su diámetro es similar al de nuestra propia galaxia, podemos calcular su distancia a partir de su tamaño aparente. Está a 30 millones de años luz (3×10^{23} metros) de la Tierra.

que subtiende en el cielo; conocemos su diámetro y calculamos su distancia. ¡Nuevamente triangulación!

Fotografías de galaxias sumamente distantes se han obtenido recientemente con el telescopio gigante de Palomar. La figura 5-8 muestra una. Se cree ahora que algunas de estas galaxias están alrededor de medio camino del límite del universo —alejadas 10^{26} metros—. ¡La mayor distancia que podemos contemplar!

5-7 Distancias pequeñas

Pensemos ahora en distancias menores. Subdividir el metro es fácil. Sin mucha dificultad podemos marcar 1 000 espacios iguales, que juntos hacen un metro. Con algo más de dificultad, pero de una manera similar (usando un buen microscopio), podemos marcar 1 000 subdivisiones iguales del milímetro, para hacer una escala en micrones (millonésimas de metro). Es difícil continuar a escalas menores, porque no podemos "ver" objetos más pequeños que la longitud de onda de la luz visible (cerca de 5×10^{-7} metros).

No necesitamos detenernos, sin embargo, en lo que no podemos ver. Con un microscopio electrónico podemos continuar el proceso, haciendo fotografías

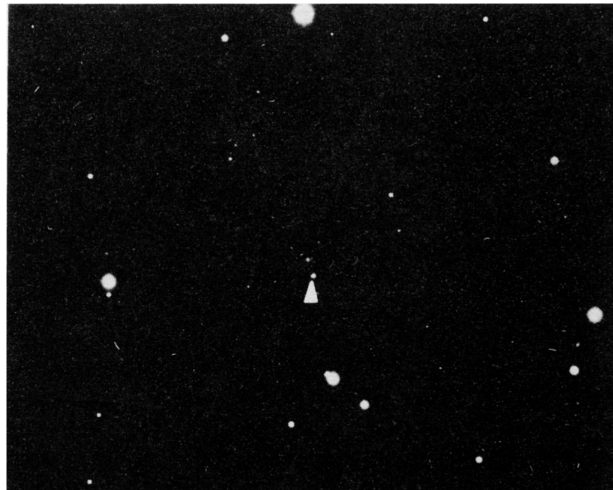


Figura 5-8. El objeto más distante a la fecha (1960). 3C295 en Boyero (señalado por la flecha), medido por el telescopio de 200 pulgadas.

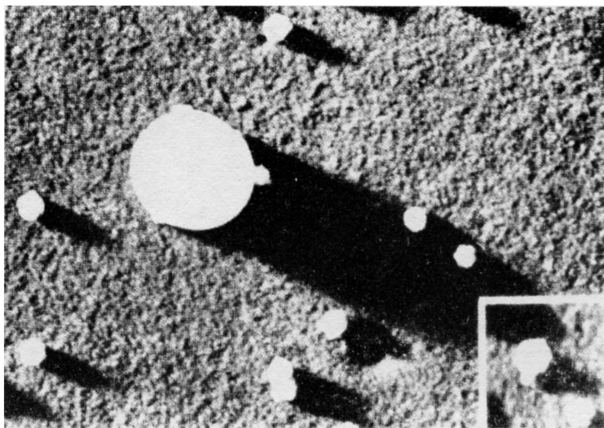


Figura 5-9. Micrografía electrónica de algunas moléculas de virus. La esfera "grande" es para la calibración y se sabe que tiene un diámetro de 2×10^{-7} metros (2 000 Å).

a una escala aún menor, digamos hasta 10^{-8} metros (figura 5-9). Mediante medidas indirectas —con una especie de triangulación a escala microscópica— podemos continuar la medida a escalas más y más pequeñas. Primero, a partir de una observación de cómo la luz de pequeña longitud de onda (radiación x) se refleja desde un conjunto de marcas de separación conocida, determinamos la longitud de onda de las vibraciones de la luz. Entonces, a partir del diagrama de dispersión de esa misma luz por un cristal, podemos determinar la posición relativa de los átomos en el cristal, obteniendo resultados que concuerdan con el espacio atómico determinado también por métodos químicos. De este modo encontramos que los átomos tienen un diámetro de alrededor de 10^{-10} metros.

Existe una gran "laguna" en los tamaños físicos entre la dimensión atómica típica de alrededor de 10^{-10} metros y las dimensiones nucleares de 10^{-15} metros, 10^{-5} veces menores. Para los tamaños nucleares, un método distinto para medir tamaños se hace conveniente. Medimos el *área aparente*, σ, llamada *sección transversal* eficaz efectiva. Si deseamos el radio, lo obtenemos de $\sigma = \pi r^2$, puesto que el núcleo es aproximadamente esférico.

Pueden hacerse medidas de la sección eficaz nuclear haciendo pasar un haz de partículas de alta energía a través de una pequeña lámina de material y observando el número de partículas que no la atraviesan. Estas partículas de alta energía se abren camino sin dificultad a través de la delgada nube de electrones y se detendrán o desviarán sólo si chocan con el peso concentrado del núcleo. Supongamos que tenemos un trozo de material de un centímetro de espesor. Habrá cerca de 10^8 capas atómicas. Pero los núcleos son tan pequeños que existe una pequeña probabilidad de que algún núcleo quede detrás de otro. Podemos *imaginar* ver la situación ampliada de manera grande —mirada a lo largo del haz de partículas— se vería como la figura 5-10.

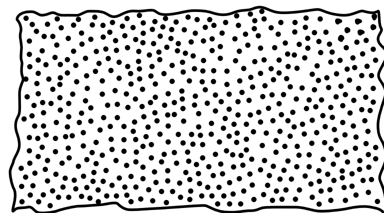


Figura 5-10. Vista imaginaria a través de un bloque de carbono de 1 cm de espesor si sólo se observaran los núcleos.

La oportunidad de que una pequeña partícula choque al núcleo en la travesía es justamente el área total cubierta por los perfiles de los núcleos, dividida entre el área total en la figura. Supongamos que sabemos que en un área A de nuestra lámina de material hay N átomos (por supuesto, uno por cada núcleo). Entonces la fracción del área "cubierta" por los núcleos es $N\sigma/A$. Ahora hagamos que el número de partículas de nuestro haz que llega a la lámina sea n_1 y el número que sale al otro lado sea n_2. La fracción que *no* atraviesa es $(n_1 - n_2)/n_1$, que justamente será igual a la fracción del área cubierta. Podemos obtener el radio del núcleo de la ecuación*

$$\pi r^2 = \sigma = \frac{A}{N}\,\frac{n_1 - n_2}{n_1}.$$

* Esta ecuación es correcta sólo si el área cubierta por los núcleos es una pequeña fracción del total, si $(n_1 - n_2)/n_1$ es mucho menor que 1. De otro modo debemos hacer una corrección por el hecho de que algunos núcleos estarán parcialmente tapados por los núcleos frente a ellos.

Distancias

Años-luz	Metros	
		? ? ? ? ? ? ? ?
	10^{27}	
		Borde del universo
10^{9}		
	10^{24}	
10^{6}		A la galaxia vecina más cercana
	10^{21}	
		Al centro de nuestra galaxia
10^{3}		
	10^{18}	
		A la estrella más cercana
1		
	10^{15}	
		Radio de la órbita de Plutón
	10^{12}	
		Al Sol
	10^{9}	
		A la Luna
	10^{6}	
		Altura de un Sputnik
	10^{3}	
		Altura de una torre de TV
	1	Altura de un niño
	10^{-3}	
		Un grano de sal
	10^{-6}	
		Un virus
	10^{-9}	
		Radio de un átomo
	10^{-12}	
	10^{-15}	Radio de un núcleo
		? ? ? ? ? ? ? ?

A partir de un experimento así encontramos que los radios de los núcleos son alrededor de uno a seis veces 10^{-15} metros. La unidad de longitud 10^{-15} metros se llama *fermi*, en honor a Enrico Fermi (1901-1954).

¿Qué encontramos si vamos a distancias más pequeñas? ¿Podemos medir distancias menores? Tales preguntas no se pueden responder aún. Se ha sugerido que el misterio aún no resuelto de las fuerzas nucleares puede ser desentrañado sólo mediante alguna modificación a nuestra idea de espacio, o medidas de tales distancias pequeñas.

Puede pensarse que sería una buena idea usar alguna longitud natural como nuestra unidad de longitud —digamos el radio de la Tierra o alguna fracción de él—. Se intentó originalmente que el metro fuera tal unidad y se definió como $(\pi/2) \times 10^{-7}$ veces el radio de la Tierra. No es conveniente ni muy exacto determinar la unidad de longitud de este modo. Por mucho tiempo se ha convenido internacionalmente que el metro sea definido como la distancia entre dos marcas en una barra guardada en un laboratorio especial en Francia. Más recientemente, se ha reconocido que esta definición no es tan precisa como para ser útil, ni tan permanente o universal como uno desearía. Se está considerando en la actualidad adoptar una nueva definición, de acuerdo con cierto número (arbitrario) de longitudes de onda de una determinada línea espectral.

Las medidas de distancia y tiempo dan resultados que dependen del observador. Dos observadores que se mueven uno con respecto al otro no medirán las mismas distancias y tiempos, cuando miden lo que parece ser la misma cosa. Las distancias e intervalos de tiempo tienen diferentes magnitudes, según el sistema coordenado (o “marco de referencia”) que se use para hacer las medidas. Estudiaremos este asunto con más detalle en un capítulo posterior.

Medidas perfectamente precisas de distancias o tiempos no están permitidas por las leyes de la naturaleza. Hemos mencionado anteriormente que los errores en una medida de la posición de un objeto deben ser al menos del tamaño de

$$\Delta x \geq \hbar/2\Delta p,$$

donde $\hbar$ es una constante fundamental pequeña llamada *constante de Planck reducida* y Δp es el error en nuestro conocimiento del momentum (masa por velocidad) del objeto, cuya posición estamos midiendo. Se mencionó, además, que la indeterminación en la medida de la posición está relacionada con la naturaleza ondulatoria de las partículas.

La relatividad del espacio y el tiempo implica que la medida del tiempo tiene también un error mínimo, de hecho está dado por

$$\Delta t \geq \hbar/2\Delta E,$$

donde ΔE es el error de nuestro conocimiento de la energía del proceso, cuyo periodo de tiempo estamos midiendo. Si deseamos saber *más* precisamente *cuándo* ocurrió algo, debemos saber menos acerca de *qué* sucedió, ya que nuestro conocimiento de la energía involucrada será menor. La indeterminación del tiempo también está relacionada con la naturaleza ondulatoria de la materia.

6

Probabilidad

"La verdadera lógica de este mundo está en el cálculo de probabilidades."
James Clerk Maxwell

6-1 Posibilidad y probabilidad

"Posibilidad" es una palabra de uso común en la vida diaria. Los reportajes de la radio al hablar del tiempo para mañana suelen decir: "hay un 60 % de posibilidad de lluvia". Ustedes suelen decir: "hay una pequeña posibilidad de que yo viva hasta los 100 años". Los científicos también usan la palabra posibilidad. Un sismólogo puede estar interesado en la pregunta: "¿Cuál es la posibilidad de que haya un terremoto de cierta magnitud al sur de California el próximo año?" Un físico puede preguntarse: "¿Cuál es la posibilidad de que cierto contador Geiger registre 20 cuentas en los próximos 10 segundos?" Un político o un hombre de Estado puede estar interesado en la pregunta: "¿Cuál es la posibilidad de que haya una guerra nuclear dentro de los próximos 10 años?" Ustedes pueden estar interesados en la posibilidad de que aprendan algo en este capítulo.

Por *posibilidad* queremos expresar algo como una conjetura. ¿Por qué planteamos conjeturas? Hacemos conjeturas cuando deseamos hacer un juicio, pero tenemos una información incompleta o un conocimiento incierto. Queremos hacer una conjetura de cómo son las cosas, o qué cosas es posible que ocurran. A menudo deseamos hacer una conjetura porque tenemos que tomar una decisión. Por ejemplo: "¿Llevaré mi impermeable mañana? ¿Contra qué movimientos de la Tierra debo diseñar un nuevo edificio? ¿Me construiré un refugio contra el polvo radiactivo? ¿Debo cambiar mi posición frente a las negociaciones internacionales? ¿Iré hoy a clases?"

En ocasiones hacemos conjeturas porque deseamos, con nuestro limitado conocimiento, decir tanto como *podamos* acerca de alguna situación. En realidad, cualquier generalización es de la naturaleza de una conjetura. Cualquier teoría física es una especie de conjetura. Hay conjeturas buenas y malas. La teoría de la probabilidad es un sistema para hacer conjeturas mejores. El lenguaje de las probabilidades nos permite hablar cuantitativamente acerca de alguna situación que puede ser altamente variable, pero que tenga un comportamiento promedio consistente.

Consideremos el lanzamiento de una moneda. Si el lanzamiento —y la moneda— son "honestos", no hay manera de saber qué esperar como resultado de un lanzamiento particular. Sin embargo, esperaríamos que con un mayor número de lanzamientos debiera haber aproximadamente igual número de caras y cruces. Decimos: "La probabilidad de que un lanzamiento resulte cara es 0.5".

Hablamos de probabilidad sólo para observaciones que contemplamos que se realicen en el futuro. *Por "probabilidad" de un resultado particular de una observación entendemos nuestra estimación de la fracción más probable de un número de observaciones repetidas que darán ese resultado particular.* Si imaginamos repetir una observación —como mirar una moneda recientemente lanzada— N veces, y si llamamos N_A *nuestra estimación* del número más probable de nuestras observaciones que dará algún resultado especificado A, por ejemplo el resultado "caras", entonces por $P(A)$, la probabilidad de observar A, entendemos

$$P(A) = N_A/N. \tag{6.1}$$

Nuestra definición requiere varios comentarios. En primer lugar, podemos hablar de probabilidad de algún suceso sólo si la ocurrencia es un resultado

posible de alguna observación *repetible*. No está claro que tenga algún sentido preguntar: “¿Cuál es la probabilidad de que haya un fantasma en esa casa?”

Ustedes pueden objetar que ninguna situación es *exactamente* repetible. Esto es cierto. Cada observación diferente debe al menos ser en distinto tiempo o lugar. Todo lo que podemos decir es que las observaciones “repetidas” deberían, para nuestros fines, *parecer equivalentes*. Al menos, supondremos que cada observación fue hecha en una situación equivalentemente preparada, y en especial con el mismo grado de ignorancia al comienzo. (¡Si en un juego de naipes miramos disimuladamente las cartas de un oponente, la estimación de nuestra posibilidad de ganar es diferente que si no lo hiciéramos!)

Debemos poner énfasis que N y N_A en la ecuación (6.1) *no* pretenden representar números basados en observaciones reales. N_A es nuestra mejor *estimación* de lo que *ocurriría* en N observaciones *imaginadas*. La probabilidad depende, por lo tanto, de nuestro conocimiento y de nuestra habilidad para hacer estimaciones. En realidad, ¡de nuestro sentido común! Afortunadamente, hay cierto grado de acuerdo en el sentido común para muchas cosas, de modo que distinta gente hará la misma estimación. Sin embargo, las probabilidades no necesitan ser números “absolutos”. Ya que dependen de nuestra ignorancia, pueden llegar a ser diferentes si nuestro conocimiento cambia.

Pueden ustedes haber notado otro aspecto bastante “subjetivo” de nuestra definición de probabilidad. Nos hemos referido a N_A como “nuestra estimación del número más probable. . . ” No queremos decir que esperamos observar *exactamente* N_A sino que esperamos un número *cercano a* N_A y que el número N_A es *más probable* que cualquier otro en la vecindad. Si lanzamos una moneda, digamos 30 veces, esperaríamos que no sería muy probable que el número de caras sea exactamente 15, sino más bien un número cercano a 15, digamos $12, 13, 14, 15, 16$ o 17. Sin embargo, si *tenemos* que elegir, decidiríamos que 15 caras es *más probable* que ningún otro número. Escribiríamos $P(\text{caras}) = 0.5$.

¿Por qué elegimos 15 como más probable que cualquier otro número? Debemos haber razonado con nosotros mismos de la siguiente manera: si el número más probable de caras es N_C en un número total de lanzamientos N, entonces el número más probable de cruces N_Z es $(N - N_C)$. (¡Estamos suponiendo que cada lanzamiento *da o bien* cara o cruz, y no “otro” resultado!) Pero si la moneda es “honesta”, no hay preferencia por caras o cruces. Hasta que tengamos alguna razón para pensar que la moneda (o el lanzamiento) es deshonesta, debemos conceder igual probabilidad a caras y cruces. De manera que escribimos $N_Z = N_C$. Por consiguiente, $N_Z = N_C = N/2$, o $P(C) = P(Z) = 0.5$.

Podemos generalizar nuestro razonamiento a *cualquier* situación en que hay m resultados posibles diferentes pero “equivalentes” (esto es, igualmente probables) de una observación. Si una observación da m resultados diferentes, y tenemos razones para creer que cualquiera de ellos es tan probable como cualquier otro, entonces la probabilidad de un resultado *particular* A es $P(A) = 1/m$.

Si hay siete bolas de diferentes colores en una caja opaca y sacamos una “al azar” (esto es, sin mirar), la probabilidad de obtener una bola de un color particular es $\frac{1}{7}$. La probabilidad de sacar a ciegas el 10 de corazones de un naipe barajado de 52 cartas es $\frac{1}{52}$. La probabilidad de lograr dos ases con dados es $\frac{1}{36}$.

En el capítulo 5 describimos el tamaño de un núcleo en términos de su área aparente, o “sección eficaz”. Cuando lo hicimos, estábamos hablando realmente, acerca de probabilidades. Cuando disparamos una partícula de alta energía contra una delgada lámina de material, hay cierta posibilidad de que pase derecho y cierta posibilidad de que dé en un núcleo. (Ya que el núcleo es tan pequeño que no podemos *verlo*, no podemos apuntar derecho al núcleo. Debemos “disparar a ciegas”.) Si hay n átomos en nuestra lámina y el núcleo de cada átomo tiene una sección eficaz de área σ, entonces el área total “sombreada” por el núcleo es $n\sigma$. Para un gran número N de disparos al azar, esperamos que el número de

impactos N_C en *algún* núcleo sea a N como el área sombreada es al área total de la lámina:

$$N_C/N = n\sigma/A. \tag{6.2}$$

Podemos decir, por lo tanto, que la *probabilidad* de que cualquier partícula proyectil sufra una colisión al pasar a través de la lámina es

$$P_C = \frac{n}{A}\,\sigma, \tag{6.3}$$

donde n/A es el número de átomos por unidad de área en nuestra lámina.

6-2 Fluctuaciones

Nos gustaría, ahora, usar nuestras ideas sobre probabilidad para considerar con más detalle la pregunta: "¿Cuántas caras puedo realmente *esperar* si lanzo N veces una moneda?" Sin embargo, antes de contestar esta pregunta, veamos lo que pasa en tal "experimento". La figura 6-1 muestra los resultados obtenidos en los primeros tres juegos cuando $N = 30$. Las secuencias de "caras" y "cruces" se muestran tal como se obtuvieron. El primer juego dio 11 caras; el segundo también dio 11; el tercero 16. En los tres experimientos ninguna vez obtuvimos 15 caras. ¿Debemos empezar a sospechar de la moneda? ¿O estábamos equivocados al pensar que el número más probable de caras en ese juego era 15? Se hicieron 97 juegos más para obtener un total de 100 experimentos de 30 lanzamientos cada uno. Los resultados de los experimentos se dan en la tabla 6-1.*

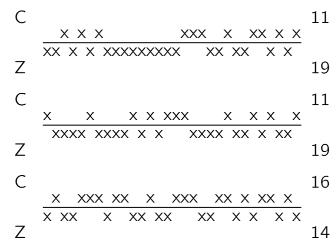


Figura 6-1. Secuencias observadas de caras y cruces en tres juegos de 30 lanzamientos cada uno.

Al mirar los números de la tabla 6-1, vemos que la mayoría de los resultados son "próximos" a 15, y que se encuentran entre 12 y 18. Podemos tener una mejor idea de los detalles de estos resultados, si hacemos un gráfico de la *distribución* de los resultados. Contamos el número de juegos en que se obtuvo el resultado k, y graficamos este número para cada k. En la figura 6-2 se muestra dicha gráfica. El resultado de 15 caras se obtuvo en 13 juegos. El resultado de 14 caras también se obtuvo 13 veces. Los resultados 16 y 17 se obtuvieron *más* de 13 veces. ¿Vamos a concluir que hay una tendencia hacia las caras? ¿No fue suficientemente buena nuestra "mejor estimación"? ¿Debemos concluir ahora que el resultado "más probable" de una serie de 30 lanzamientos es en realidad 16 caras? ¡Esperen! En todos los juegos tomados en conjunto hubo 3 000 lanzamientos. Y el número total de caras obtenidas fue 1 493. La fracción de lanzamientos que dieron caras es 0.498, muy cercana, pero ligeramente *menor* que la mitad. ¡Por cierto, *no* podríamos suponer que la probabilidad de obtener cara sea mayor que 0.5! El hecho que un conjunto *particular* de observaciones dio 16 caras más a menudo es una *fluctuación*. Aún esperamos que el número *más probable* de caras sea 15.

Tabla 6-1

Número de caras en ensayos sucesivos de 30 lanzamientos de una moneda.

11	16	17	15	17	16	19	18	15	13
11	17	17	12	20	23	11	16	17	14
16	12	15	10	18	17	13	15	14	15
16	12	11	22	12	20	12	15	16	12
16	10	15	13	14	16	15	16	13	18
14	14	13	16	15	19	21	14	12	15
16	11	16	14	17	14	11	16	17	16
19	15	14	12	18	15	14	21	11	16
17	17	12	13	14	17	9	13	19	13
14	12	15	17	14	10	17	17	12	11

} 100 juegos

* Después de los tres primeros juegos, el experimento fue realmente hecho metiendo 30 monedas dentro de una caja y agitándola fuertemente, luego se contaron el número de caras que resultaron.

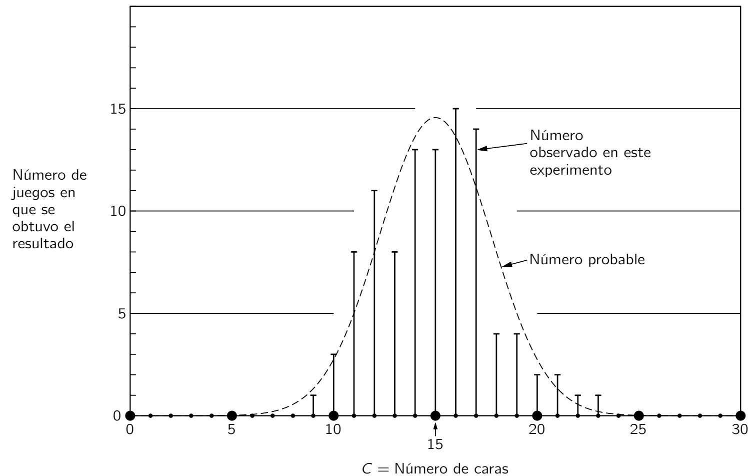


Figura 6-2. Resumen de los resultados de 100 juegos de 30 lanzamientos cada uno. Las barras verticales indican el número de juegos en los cuales se obtuvo un resultado de *C* caras. La curva punteada muestra los números esperados calculados con probabilidad.

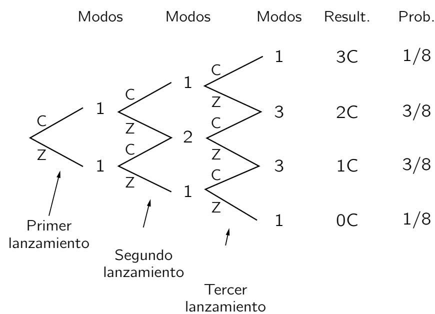


Figura 6-3. Un diagrama para demostrar el número de modos en que se puede obtener un resultado de 0, 1, 2 o 3 cara en un juego de tres lanzamientos.

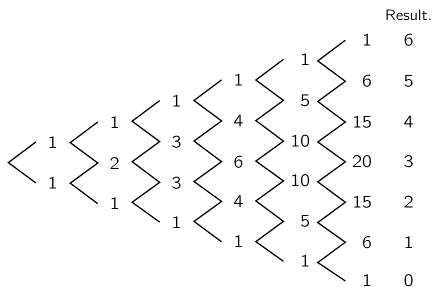


Figura 6-4. Un diagrama semejante al de la figura 6-3, para un juego de seis lanzamientos.

Podemos hacer la pregunta: "¿Cuál *es* la probabilidad de que en un juego de 30 lanzamientos resulten 15 caras, o 16, o cualquier otro número?" Hemos dicho que en el experimento de un lanzamiento, la probabilidad de obtener *una* cara es 0.5 y que la probabilidad de no obtener cara es 0.5. En un experimento de dos lanzamientos hay *cuatro* resultados posibles: *cara-cara, cara-cruz, cruz-cara, cruz-cruz.* Ya que cada una de estas secuencias es igualmente probable, concluimos que (a) la probabilidad para lograr dos caras es $\frac{1}{4}$, (b) la probabilidad de obtener una cara es $\frac{2}{4}$, (c) la probabilidad de no obtener ninguna es $\frac{1}{4}$. Hay *dos* modos de obtener una cara, pero sólo uno para obtener dos caras o no obtener caras.

Consideren ahora un juego de tres lanzamientos. En el tercer lanzamiento es igualmente probable que resulten caras o cruces. Hay un solo modo de obtener tres caras: *debemos* haber obtenido dos caras en los dos primeros lanzamientos, y cara en el tercero. Hay, sin embargo, *tres* modos de obtener dos caras. Podríamos tirar cruz después de haber tirado dos caras (un modo) o podríamos tirar cara después de tirar sólo una cara en los dos primeros lanzamientos (dos modos). De manera que para los resultados 3-*caras*, 2-*caras*, 1-*cara*, 0-*cara* tenemos que el número de modos igualmente posibles es $1, 3, 3, 1$, con un total posible de ocho secuencias diferentes. Las probabilidades son $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, $\frac{1}{8}$.

Los razonamientos que hemos estado haciendo pueden resumirse en un diagrama como el de la figura 6-3. Está claro cómo puede continuarse el diagrama para juegos con un número mayor de lanzamientos. La figura 6-4 muestra dicho diagrama para un juego de seis lanzamientos. El número de "modos" en cualquier punto del diagrama es justamente el número de "caminos" diferentes (secuencias de caras y cruces) que pueden tomarse desde el punto de partida. La columna vertical nos da el número total de caras logradas. El conjunto de números que aparece en dicho diagrama se conoce como *triángulo de Pascal.* Los números se conocen también como *coeficientes binomiales*, porque también aparecen en el desarrollo de $(a+b)^n$. Si llamamos n al número de lanzamientos y k al número de caras lanzadas, entonces los números del diagrama se designan usualmente con el símbolo $\binom{n}{k}$. Podemos escribir el paso de cómo los coeficientes binomiales se pueden calcular:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \tag{6.4}$$

donde $n!$ es llamado el "factorial de n" y representa el siguiente producto:

$$(n)(n-1)(n-2)\cdots(3)(2)(1).$$

Ahora estamos listos para calcular la probabilidad $P(k,n)$ cuando se obtienen k caras en n lanzamientos, usando nuestra definición, ecuación (6.1). El número total de secuencias posibles es 2^n (ya que hay dos resultados para cada lanzamiento), y el número de maneras de obtener k caras es $\binom{n}{k}$, todas igualmente probables, de modo que tenemos:

$$P(k,n) = \frac{\binom{n}{k}}{2^n}. \tag{6.5}$$

Ya que $P(k,n)$ es la fracción que representa el número de juegos que consideramos que den k caras, entonces esperaríamos encontrar en 100 juegos el número k de caras que está dado por $100 \cdot P(k,n)$ veces. La curva punteada de la figura 6-2 pasa por los puntos calculados con $100 \cdot P(k,30)$. Vemos que *esperamos* obtener un resultado de 15 caras en 14 o 15 juegos, mientras que este resultado fue observado en 13 juegos. *Esperamos* un resultado de 16 en 13 o 14 juegos, pero obtuvimos dicho resultado en 16 juegos. Tales fluctuaciones son "parte del juego".

El método que acabamos de usar puede aplicarse a la situación más general en que sólo hay dos posibles resultados de una simple observación. Designemos los dos resultados por G (por "gana") y P (por "pierde"). En el caso general, la probabilidad de G o de P en un evento simple no necesita ser igual. Sea p la probabilidad de obtener el resultado G. Entonces q, la probabilidad para P, es necesariamente $(1-p)$. En un conjunto de n ensayos, la probabilidad $P(k,n)$ de que G se obtenga k veces es:

$$P(k,n) = \tbinom{n}{k} p^k q^{n-k}. \tag{6.6}$$

Esta función de probabilidad se llama probabilidad de *Bernouilli*, o también se le denomina *binomial*.

6-3 La caminata aleatoria

Hay otro interesante problema en el cual se requiere la idea de probabilidad. Es el problema de la "caminata aleatoria". En su versión más simple, imaginamos un "juego" en que un "jugador" parte del punto $x = 0$, y en cada "movimiento" se requiere que dé un paso, *ya sea* hacia adelante (hacia $+x$) *o* hacia atrás (hacia $-x$). La elección debe hacerse *al azar*, determinada, por ejemplo, por el lanzamiento de una moneda. ¿Cómo describiremos el movimiento resultante? En su forma general el problema está relacionado con el movimiento de los átomos (u otras partículas) en un gas —llamado movimiento browniano— y también a la combinación de errores en las medidas. Ustedes verán que el problema de la caminata aleatoria está estrechamente relacionado con el problema del lanzamiento de una moneda que ya hemos discutido.

En primer lugar, veamos unos pocos ejemplos de una caminata aleatoria. Podemos caracterizar el progreso del caminante por la distancia neta D_N recorrida en N pasos. En el gráfico de la figura 6-5 mostramos tres ejemplos de la trayectoria de un caminante al azar. (Hemos usado para la secuencia al azar de las elecciones los resultados de los lanzamientos de monedas que se muestran en la figura 6-1.)

¿Qué podemos decir de tal movimiento? En primer lugar podemos preguntar: "¿Hasta dónde irá en promedio?" Debemos *esperar* que su progreso promedio será cero, ya que es igualmente probable ir hacia adelante o hacia atrás. Pero tenemos la impresión de que al crecer N, es más probable que se haya apartado del punto de partida. Por lo tanto, podemos preguntar cuál es la distancia media viajada en *valor absoluto*, es decir, cuál es el promedio de $|D|$. Sin embargo, es más conveniente tratar con otra medida del "progreso", el cuadrado de la distancia: D^2 es positivo, ya sea para el movimiento positivo o negativo, y por lo tanto es una razonable *medida* de dicho vagabundeo al azar.

Podemos demostrar que el valor esperado para D_N^2 es justamente N, el número de pasos dados. Por "valor esperado" queremos indicar el valor probable (nuestra mejor estimación), que podemos imaginar como el comportamiento promedio

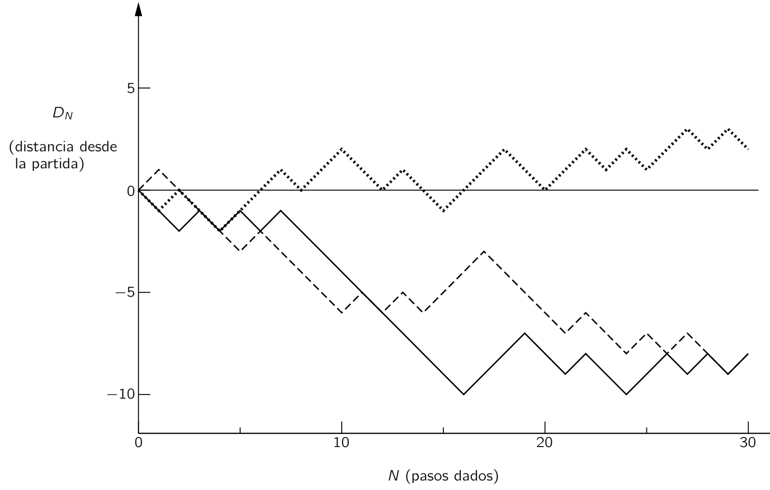


Figura 6-5. El progreso realizado en una caminata aleatoria. La coordenada horizontal N es el número total de pasos dados; la coordenada vertical D_N es la distancia neta desplazada desde la posición inicial.

esperado de *muchas* secuencias *repetidas*. Representamos ese valor esperado por $\langle D_N^2\rangle$, y podemos referirnos a él como la “media del cuadrado de la distancia”. Después de un paso, D^2 es siempre $+1$, de manera que tenemos por cierto $\langle D_1^2\rangle = 1$. (Todas las distancias estarán medidas en términos de la unidad de un paso. No continuaremos escribiendo las unidades de distancia.)

El valor esperado de D_N^2 para $N > 1$ puede obtenerse a partir de D_{N-1}. Si después de $(N-1)$ pasos tenemos D_{N-1}, entonces después de N pasos tenemos $D_N = D_{N-1} + 1$ o $D_N = D_{N-1} - 1$. Para los cuadrados,

$$D_N^2 = \begin{cases} D_{N-1}^2 + 2D_{N-1} + 1, \\ \qquad\quad o \\ D_{N-1}^2 - 2D_{N-1} + 1. \end{cases} \tag{6.7}$$

En cierto número de secuencias independientes, esperamos obtener en cada valor la mitad de las veces, así nuestra expectativa promedio es justamente el promedio de los dos valores posibles. El valor esperado para D_N^2 es entonces $D_{N-1}^2 + 1$. *En general*, podríamos esperar para D_{N-1}^2 su “valor esperado” $\langle D_{N-1}^2\rangle$ (¡por definición!). Así,

$$\langle D_N^2\rangle = \langle D_{N-1}^2\rangle + 1. \tag{6.8}$$

Ya hemos demostrado que $\langle D_1^2\rangle = 1$, se concluye entonces que:

$$\langle D_N^2\rangle = N, \tag{6.9}$$

¡un resultado particularmente fácil!

Si deseamos un número tal como una distancia, en lugar de una distancia al cuadrado, para representar el “progreso hecho más *allá* del origen” en el movimiento casual, podemos usar la “distancia media cuadrática” D_{mc}:

$$D_{\text{mc}} = \sqrt{\langle D^2\rangle} = \sqrt{N}. \tag{6.10}$$

Hemos indicado que la matemática de la caminata aleatoria es muy similar a la del juego del lanzamiento de la moneda que consideramos al comienzo de este capítulo. Si imaginamos que la dirección de cada paso está en correspondencia con la aparición de caras o cruces en un lanzamiento de moneda, entonces D es justamente $N_C - N_Z$, la diferencia en el número de caras y cruces. Ya que $N_C + N_Z = N$, el número total de pasos (y de lanzamientos), tenemos

$D = 2N_C - N$. Hemos deducido antes una expresión para la distribución esperada de N_C (llamada también k) y obtenido el resultado de la ecuación (6.5). Ya que N es justamente una constante, tenemos la distribución correspondiente para D. (Ya que por cada cara por sobre $N/2$ hay una cruz "ausente", tenemos el factor 2 entre N_C y D.) La gráfica de la figura 6-2 representa la distribución de las distancias que podemos obtener en 30 pasos al azar (donde $k = 15$ debe leerse $D = 0$; $k = 16$, $D = 2$; etcétera).

La variación de N_C de su valor esperado $N/2$ es:

$$N_C - \frac{N}{2} = \frac{D}{2}. \tag{6.11}$$

La desviación media cuadrática es:

$$\left(N_C - \frac{N}{2}\right)_{\text{mc}} = \tfrac{1}{2}\sqrt{N}. \tag{6.12}$$

De acuerdo con nuestro resultado para D_{mc}, esperamos que la distancia "típica" en 30 pasos debería ser: $\sqrt{30} \approx 5.5$, o para un k típico sería alrededor de $5.5 \div 2 = 2.75$ unidades a partir de 15. Vemos que el "ancho" de la curva en la figura 6-2, medido desde el centro, es justamente de alrededor de 3 unidades, en concordancia con este resultado.

Estamos ahora en situación de considerar una pregunta que hemos evitado hasta ahora. ¿Cómo podremos decir si una moneda es "honesta" o está "cargada"? Al menos, podemos dar ahora una respuesta parcial. Para una moneda honesta, esperamos que la fracción de veces que aparecen caras sea 0.5, es decir,

$$\frac{\langle N_C \rangle}{N} = 0.5. \tag{6.13}$$

También esperamos que un N_C real se desvíe de $N/2$ en aproximadamente $\sqrt{N}/2$, o que la *fracción* se desvíe en:

$$\frac{1}{N}\,\frac{\sqrt{N}}{2} = \frac{1}{2\sqrt{N}}.$$

Cuanto más grande es N, entonces esperamos que esté más cerca de la fracción N_C/N a un medio.

En la figura 6-6 hemos puesto la fracción N_C/N para los lanzamientos de moneda señalados al comienzo de este capítulo. Vemos que la tendencia del número de caras se aproxima a 0.5 cuando N es grande. Desafortunadamente, para cualquier lanzamiento, o combinación de juegos, no hay *garantía* que la desviación observada esté siquiera *cerca* de la desviación *esperada*. Existe siempre la posibilidad finita que una gran fluctuación —una larga ristra de caras o de cruces— dé una desviación arbitrariamente grande. Todo lo que podemos decir es que *si* la desviación es cercana al valor esperado $1/2\sqrt{N}$ (digamos dentro de un factor 2 o 3), no tenemos razón para sospechar de la honestidad de la moneda. Si es mucho mayor, podríamos sospechar, pero no probar que la moneda está cargada (¡o que el que la lanza sea listo!).

Tampoco hemos considerado cómo tratar el caso de una "moneda" o de algún otro objeto de "suerte" similar (por ejemplo, una piedra que siempre

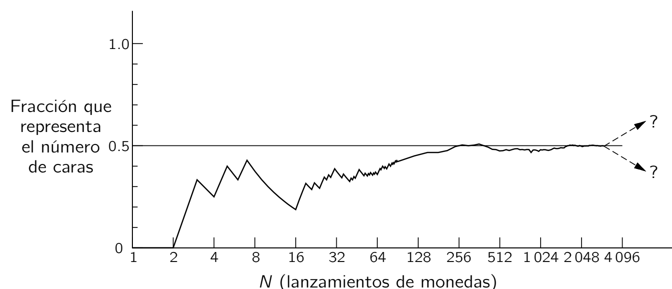


Figura 6-6. La fracción de los lanzamientos que dio caras en una secuencia particular de N lanzamientos de una moneda.

cae en una de dos posiciones) tal que tengamos una buena razón para creer que *tenga* una probabilidad diferente para caras o cruces. Hemos definido $P(C) = \langle N_C \rangle / N$. ¿Cómo sabremos qué *esperar* para N_C? En algunos casos, lo mejor que podemos hacer es observar el número de caras obtenidas en un gran número de lanzamientos. A falta de algo mejor, debemos poner $\langle N_C \rangle = N_C$ (observado). (¿Cómo podríamos esperar algo más?) Debemos comprender, sin embargo, que en tal caso un experimento diferente, o un observador diferente, podría concluir que $P(C)$ era diferente. Sin embargo, deberíamos *esperar* que las diferentes respuestas concuerdan con la desviación $1/2\sqrt{N}$ [si $P(C)$ es cercano a un medio]. Un físico experimental generalmente dice que una probabilidad "experimentalmente determinada" tiene un "error", y escribe:

$$P(C) = \frac{N_C}{N} \pm \frac{1}{2\sqrt{N}}. \tag{6.14}$$

Hay una implicación en esta expresión que dice que *existe* una probabilidad "verdadera" o "correcta" que *podría* calcularse si supiéramos bastante, y que la observación puede tener "error" debido a una fluctuación. Sin embargo, no hay modo de hacer tal razonamiento lógicamente consistente. Tal vez es mejor reconocer que el concepto de probabilidad se considera en sentido subjetivo, que está siempre basado en un conocimiento incierto, y que su evaluación cuantitativa está sujeta a cambio a medida que obtenemos más información.

6-4 Una distribución de probabilidad

Volvamos ahora a la caminata aleatoria y consideremos una modificación de ella. Supongamos que, además de una elección al azar de la *dirección* (+ o −) de cada paso, la *longitud* de cada paso también varía de un modo impredecible, siendo la única condición de que *en promedio* la longitud del paso sea una unidad. Este caso es más representativo de algo como el movimiento térmico de una molécula en un gas. Si llamamos S a la longitud de un paso, entonces S puede tener cualquier valor, pero lo más común es que esté "cercano" a 1. Para ser específicos, hagamos $\langle S^2 \rangle = 1$ o, igualmente, $S_{\text{mc}} = 1$. Nuestra deducción para $\langle D^2 \rangle$ proseguiría como antes excepto que la ecuación (6.8) deberá ahora cambiarse y se leerá

$$\langle D_N^2 \rangle = \langle D_{N-1}^2 \rangle + \langle S^2 \rangle = \langle D_{N-1}^2 \rangle + 1. \tag{6.15}$$

Tenemos, como antes, que:

$$\langle D_N^2 \rangle = N. \tag{6.16}$$

¿Qué esperaríamos ahora para la distribución de distancias D? ¿Cuál es, por ejemplo, la probabilidad para $D = 0$ después de 30 pasos? ¡La respuesta es cero! La probabilidad de que D sea *cualquier* valor *particular* es cero, ya que no hay posibilidad alguna que la suma de todos los pasos hacia atrás (de longitudes variables) sea exactamente igual a la suma de los pasos hacia adelante. No podemos hacer una gráfica como la de la figura 6-2.

Sin embargo, podemos obtener una representación similar a la de la figura 6-2 si preguntamos, no por la probabilidad de obtener D exactamente igual a 0, 1 o 2, sino en cambio cuál es la probabilidad de obtener D *cercano* a 0, 1 o 2. Definamos $P(x, \Delta x)$ como la probabilidad de que D caiga en el intervalo Δx ubicado en x (digamos de x a $x + \Delta x$). Esperamos que para Δx pequeño, la probabilidad de que D caiga en dicho intervalo sea proporcional a Δx, el ancho del intervalo. Así podemos escribir:

$$P(x, \Delta x) = p(x)\, \Delta x. \tag{6.17}$$

La función $p(x)$ se llama *densidad de probabilidad*.

La forma de $p(x)$ dependerá de N, el número de pasos dados, y también de la distribución de las longitudes de los pasos individuales. No podemos demostrarlo

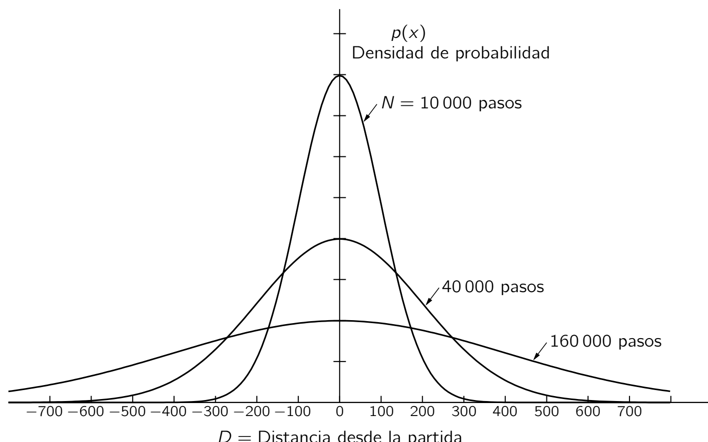


Figura 6-7. La densidad de probabilidad para alcanzar la distancia D desde el lugar de partida en una caminata aleatoria de N pasos. (D se mide en unidades del valor medio cuadrático de la longitud del paso.)

aquí, pero para N grande, $p(x)$ es la *misma* para todas las distribuciones razonables de longitudes de pasos individuales, y depende sólo de N. En la figura 6-7 graficamos $p(x)$ para tres valores de N. Notarán que el "ancho medio" (desviación típica a partir de $x = 0$) de estas curvas es $\sqrt{N}$, como hemos demostrado que debe ser.

Además, ustedes pueden notar que el valor de $p(x)$ cerca de cero es inversamente proporcional a $\sqrt{N}$. Esto sucede porque las curvas son todas de la misma forma y sus áreas bajo las curvas deben ser todas iguales. Ya que $p(x)\Delta x$ es la probabilidad de encontrar D en Δx cuando Δx es pequeño, podemos determinar la probabilidad de encontrar D en *alguna parte* dentro de un intervalo arbitrario desde x_1 a x_2, dividiendo el intervalo en un número de pequeños incrementos Δx y evaluando la suma de los términos $p(x)\Delta x$ para cada incremento. La probabilidad de que D caiga en alguna parte entre x_1 y x_2, que podemos escribir $P(x_1 < D < x_2)$, es igual al área sombreada en la figura 6-8. Mientras más pequeño tomemos el incremento Δx más correcto es nuestro resultado. Por lo tanto, podemos escribir:

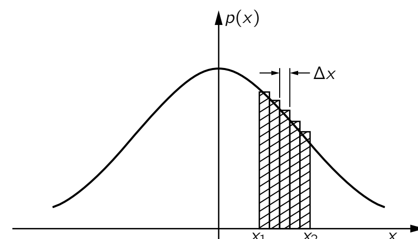


Figura 6-8. La probabilidad de que la distancia D recorrida en una caminata aleatoria esté comprendida entre x_1 y x_2 es la superficie bajo la curva de $p(x)$ de x_1 a x_2.

$$P(x_1 < D < x_2) = \sum p(x)\,\Delta x = \int_{x_1}^{x_2} p(x)\,dx. \tag{6.18}$$

El área bajo la curva completa es la probabilidad de que D caiga en alguna parte (esto es, que tenga *algún* valor entre $x = -\infty$ y $x = +\infty$). Esa probabilidad es ciertamente 1. Debemos tener que:

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1. \tag{6.19}$$

Ya que las curvas de la figura 6-7 se hacen más anchas en proporción a $\sqrt{N}$, sus alturas deben ser proporcionales a $1/\sqrt{N}$ para mantener el área total igual a 1.

La función densidad de probabilidad que hemos descrito es una de las que se encuentran más comúnmente. Se conoce como la densidad de probabilidad *normal* o *gaussiana.* Tiene la forma matemática

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\,e^{-x^2/2\sigma^2}, \tag{6.20}$$

donde σ se llama *desviación normal* y en nuestro caso está dada por $\sigma = \sqrt{N}$, o si la raíz de la magnitud del paso medio cuadrático es diferente de 1 por $\sigma = \sqrt{N}S_{\text{mc}}$.

Indicamos antes que el movimiento de una molécula, o de cualquier partícula, en un gas es como una caminata aleatoria. Supongamos que abrimos una botella de un compuesto orgánico y dejamos que algo de su vapor escape al aire. Si hay

corrientes de aire, de manera que el aire esté circulando, las corrientes llevarán también consigo el vapor. Pero aun en *aire perfectamente quieto*, el vapor gradualmente se esparcirá —difundirá— hasta que haya penetrado enteramente en la sala. Podemos detectarlo por su color o su olor. Las moléculas individuales del vapor orgánico se esparcen en el aire quieto debido a los movimientos moleculares causados por colisiones con otras moléculas. Si conocemos el tamaño del "paso" medio y el número de pasos que dan por segundo, podemos encontrar la probabilidad de que una, o varias moléculas, sean encontradas a cierta distancia de su punto de partida después de cualquier intervalo de tiempo. A medida que pasa el tiempo, más pasos se dan y el gas se esparce como en las curvas sucesivas de la figura 6-7. En un capítulo posterior encontraremos que el tamaño de los pasos y sus frecuencias están relacionados con la temperatura y presión del gas.

Dijimos antes que la presión de un gas se debe a que las moléculas rebotan contra las paredes del recipiente. Cuando lleguemos más tarde a hacer una descripción más cuantitativa, desearemos saber la rapidez con que se mueven las moléculas cuando rebotan, ya que el impacto que hacen dependerá de esa velocidad. Sin embargo, no podemos hablar de *la* velocidad de las moléculas. Es necesario usar una descripción probabilística. Una molécula puede tener cualquier velocidad, pero algunas velocidades son más probables que otras. Describimos lo que está ocurriendo diciendo que la probabilidad de que cualquier molécula particular tenga una velocidad entre v y $v + \Delta v$ es $p(v)\delta v$, donde $p(v)$, una densidad de probabilidad, es una función dada de la velocidad v. Veremos más tarde cómo Maxwell, usando sentido común y las ideas de probabilidad, pudo encontrar una expresión matemática para $p(v)$. La forma* de la función $p(v)$ se muestra en la figura 6-9. Las velocidades pueden tener cualquier valor, pero lo más posible es que estén cerca del valor más probable o esperado v_p.

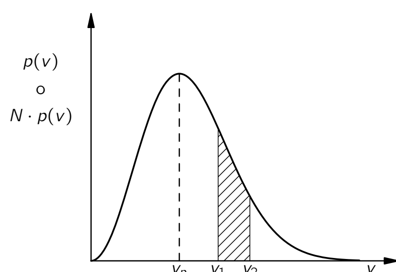


Figura 6-9. La distribución de velocidades de las moléculas en un gas.

A menudo interpretamos la curva de la figura 6-9 de un modo algo diferente. Si consideramos las moléculas en un recipiente típico (digamos con un volumen de un litro), entonces hay un número muy grande de moléculas presentes ($N \approx 10^{22}$). Ya que $p(v)\Delta v$ es la probabilidad de que *una* molécula tenga su velocidad en Δv, con nuestra definición de probabilidad queremos decir que el número *esperado* $\langle \Delta N \rangle$ que se encuentra con una velocidad en el intervalo δv está dado por

$$\langle \Delta N \rangle = N\, p(v)\, \Delta v. \tag{6.21}$$

Llamamos a $Np(v)$ "distribución de velocidad". El área bajo la curva entre dos velocidades v_1 y v_2, por ejemplo el área sombreada en la figura 6-9, representa, para la curva $Np(v)$, el número esperado de moléculas con velocidades entre v_1 y v_2. Ya que con un gas estamos tratando corrientemente con un gran número de moléculas, esperamos que las desviaciones de los números esperados sean pequeñas (como $1/\sqrt{N}$), así que, a menudo, omitimos decir el número "esperado" y decimos, en cambio: "El número de moléculas con velocidades entre v_1 y v_2 es el área bajo la curva". Sin embargo, debemos recordar que tales afirmaciones son siempre acerca de números *probables*.

6-5 El principio de incertidumbre

Las ideas de probabilidad son ciertamente útiles para describir el comportamiento de algo así como 10^{22} moléculas en una muestra de gas, porque es claramente impracticable siquiera tratar de escribir la posición y velocidad para cada molécula. Cuando la probabilidad se aplicó por primera vez a tales problemas se consideró que era una *conveniencia* —una manera de tratar situaciones muy complejas—. Ahora creemos que las ideas de probabilidad son *esenciales* para la descripción del acontecer atómico. De acuerdo con la mecánica cuántica, la teoría matemática de las partículas, hay siempre cierta incertidumbre en la *especificación* de posiciones y velocidades. A lo más podemos decir que hay

* La expresión de Maxwell es $p(v) = Cv^2 e^{-av^2}$, donde a es una constante relacionada con la temperatura y C se elige de modo que la probabilidad total sea uno.

cierta probabilidad de que cualquier partícula tenga una posición cercana a cierta coordenada x.

Podemos dar una densidad de probabilidad $p_1(x)$, tal que $p_1(x)\Delta x$ sea la probabilidad de que la partícula se encuentre entre x y $x+\Delta x$. Si la partícula está razonablemente bien localizada, digamos cerca de x_0, la función $p_1(x)$ puede estar dada por la gráfica de la figura 6-10(a). En forma similar debemos especificar la velocidad de la partícula mediante una densidad de probabilidad $p_2(v)$ con $p_2(v)\Delta v$ la probabilidad de que se encuentre la velocidad entre v y $v+\Delta v$.

Es uno de los resultados fundamentales de la mecánica cuántica que las dos funciones $p_1(x)$ y $p_2(v)$ no pueden elegirse independientemente y que, en particular, ambas no pueden hacerse arbitrariamente estrechas. Si llamamos $[\Delta x]$ el "ancho" típico de la curva $p_1(x)$, y $[\Delta v]$ al de la curva $p_2(v)$ (como se indica en la figura), la naturaleza requiere que el *producto* de ambos anchos sea al menos del tamaño del número $\hbar/2m$, donde m es la masa de la partícula. Podemos escribir esta relación básica como:

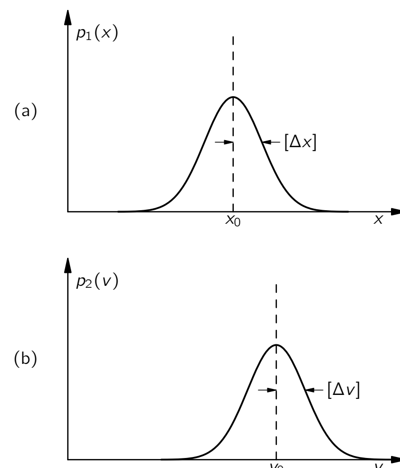


Figura 6-10. Densidades de probabilidad para la observación de la posición y velocidad de una partícula.

$$[\Delta x]\cdot[\Delta v]\geq \hbar/2m. \tag{6.22}$$

Esta ecuación es un enunciado del *principio de incertidumbre de Heisenberg* que hemos mencionado anteriormente.

Ya que el segundo miembro de la ecuación (6.22) es una constante, esta ecuación dice que si tratamos de "inmovilizar" una partícula forzándola a estar en cierto lugar, ella acaba por tener una gran velocidad. O si tratamos de forzarla a moverse muy lentamente o a una velocidad precisa, se "esparce" de modo que no sabemos muy bien justamente dónde está. ¡Las partículas se comportan de un modo divertido!

El principio de incertidumbre describe una imprecisión inherente que debe existir en cualquier intento para describir la naturaleza. Nuestra descripción más precisa de la naturaleza *debe* ser en términos de *probabilidades*. Hay algunas personas a las que no les gusta esta manera de describir la naturaleza. Creen que si pudieran decir lo que *realmente* sucede con una partícula, podrían conocer su velocidad y posición simultáneamente. En los primeros días del desarrollo de la mecánica cuántica, Einstein estuvo muy preocupado con este problema. Solía sacudir la cabeza y decir: "¡Pero, seguro que Dios no echa los dados para determinar cómo deberían moverse los electrones!" Se preocupó de ese problema largo tiempo y seguramente nunca se resignó realmente al hecho de que ésta es la mejor descripción de la naturaleza que uno puede dar. Existen todavía uno o dos físicos que trabajan en el problema, que tienen una convicción intuitiva de que es posible describir el mundo en una forma diferente y que toda esta incertidumbre acerca de cómo son las cosas puede ser eliminada. Ninguno ha tenido éxito aún.

La necesaria incertidumbre en nuestra especificación de la posición de una partícula cobra mayor importancia cuando deseamos describir la estructura de los átomos. En el átomo de hidrógeno, que tiene un núcleo con únicamente un protón y con un electrón fuera del núcleo, ¡la incertidumbre en la posición del electrón es tan grande como el átomo mismo! Por lo tanto, no podemos hablar propiamente del electrón moviéndose en alguna "órbita" alrededor del protón. Lo más que podemos decir es que hay cierta *posibilidad* $p(r)\Delta V$ de observar el electrón como un elemento del volumen ΔV a distancia r del protón. La densidad de probabilidad $p(r)$ está dada por la mecánica cuántica. Para un átomo de hidrógeno no perturbado $p(r)=Ae^{-2r/a}$, es una función con forma de campana como la de la figura 6-8. El número a es el radio "típico", donde la función decrece rápidamente. Ya que hay una pequeña probabilidad de encontrar el electrón a distancias mayores que a a partir del núcleo, podemos pensar que a es "el radio del átomo", alrededor de 10^{-10} metros.

Podemos formarnos una imagen del átomo de hidrógeno, visualizando una "nube" cuya densidad es proporcional a la densidad de probabilidad de observar el electrón. Una muestra de una nube tal se indica en la figura 6-11. De modo que nuestro mejor "retrato" de un átomo de hidrógeno es un núcleo rodeado de una "nube electrónica" (aunque *realmente* queremos indicar una "nube de

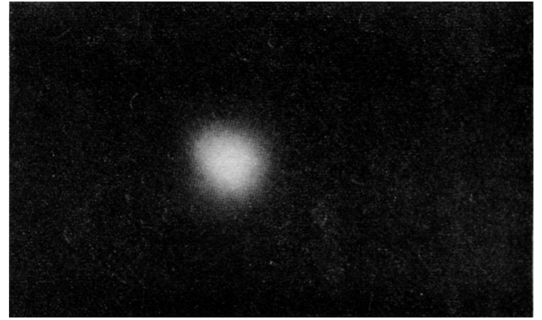


Figura 6-11. Una manera de visualizar un átomo de hidrógeno. La densidad (blancura) de la nube representa la densidad de probabilidad de observar el electrón.

probabilidad"). El electrón está allí en alguna parte, pero la naturaleza sólo nos permite conocer la probabilidad de encontrarlo en algún lugar particular.

En sus esfuerzos por aprender lo más posible acerca de la naturaleza, la física moderna ha encontrado que ciertas cosas no pueden nunca ser "conocidas" con certeza. Muchos de nuestros conocimientos deben permanecer siempre inciertos. Lo *más* que podemos saber está en términos de probabilidades.

7

La teoría de la gravitación

7-1 Movimientos planetarios

En este capítulo discutiremos una de las más amplias generalizaciones de la mente humana. Mientras admiramos la mente humana deberíamos tomar algún tiempo para venerar una *naturaleza* que pudo lograr en una forma tan acabada, y con tal generalidad, un principio tan elegantemente simple como la ley de la gravitación. ¿Qué es esta ley de la gravitación? Consiste en que todo objeto en el universo atrae a todo otro objeto con una fuerza que para dos cuerpos cualesquiera es proporcional a la masa de cada uno y varía inversamente con el cuadrado de la distancia entre ellos. Este enunciado puede expresarse matemáticamente por la ecuación

$$F = G\,\frac{mm'}{r^2}.$$

Si a esto agregamos el hecho de que un objeto responde a una fuerza acelerando en la dirección de la fuerza en una cantidad que es inversamente proporcional a la masa del objeto, habremos dicho todo lo necesario para que un matemático suficientemente talentoso pueda deducir entonces todas las consecuencias de estos dos principios. Sin embargo, puesto que no se supone todavía que ustedes sean suficientemente talentosos, discutiremos las consecuencias con más detalle y no los dejaremos meramente con estos dos principios escuetos. Relataremos brevemente la historia del descubrimiento de la ley de la gravitación y discutiremos algunas de sus consecuencias, sus efectos sobre la historia, los misterios que tal ley acarrea y algunos refinamientos de la ley hechos por Einstein; discutiremos también las relaciones de la ley con otras leyes de la física. Todo esto no puede hacerse en un capítulo, pero estos aspectos se tratarán a su debido tiempo en los capítulos siguientes.

La historia comienza con los antiguos observando el movimiento de los planetas entre las estrellas y deduciendo finalmente que ellos se movían alrededor del Sol, un hecho que más tarde fue redescubierto por Copérnico. Tomó un poco más de trabajo descubrir exactamente *cómo* los planetas se movían alrededor del Sol y exactamente con *qué movimiento*. En los comienzos del siglo XV hubo grandes debates sobre si ellos realmente se movían alrededor del Sol o no. Tycho Brahe tuvo una idea que fue diferente de cualquiera de las propuestas por los antiguos: su idea fue que estos debates acerca de la naturaleza de los movimientos de los planetas se resolverían mejor si las posiciones reales de los planetas en el cielo se midieran con suficiente precisión. Si las medidas mostraran cómo se mueven exactamente los planetas, entonces tal vez sería posible establecer uno u otro punto de vista. Ésta fue una idea tremenda —que para descubrir algo es mejor realizar algunos experimentos cuidadosos que continuar con profundos argumentos filosóficos—. Con esta idea, Tycho Brahe estudió las posiciones de los planetas durante muchos años en su observatorio de la isla de Ven, cerca de Copenhague. Realizó voluminosas tablas, que fueron estudiadas por el matemático Kepler, después de la muerte de Tycho. Kepler descubrió a partir de los datos algunas leyes muy bellas y notables, pero simples, sobre el movimiento planetario.

7-2 Leyes de Kepler

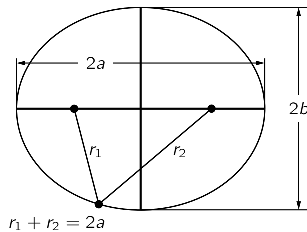


Figura 7-1. Una elipse.

En primer lugar, Kepler encontró que cada planeta se mueve alrededor del Sol en una curva llamada *elipse*, con el Sol en un foco de la elipse. Una elipse no es precisamente un óvalo, sino una curva muy específica y precisa que puede obtenerse usando dos tachuelas, una en cada foco, un lazo de cuerda y un lápiz; más matemáticamente es el lugar geométrico de todos los puntos cuya suma de las distancias a dos puntos fijos (los focos) es constante. O bien, si lo prefieren, es un círculo achatado (figura 7-1).

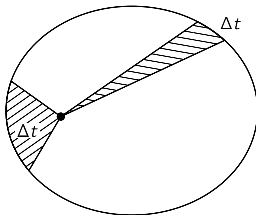


Figura 7-2. Ley de las áreas de Kepler.

La segunda observación de Kepler fue que los planetas no se mueven alrededor del Sol con velocidad uniforme, sino que se mueven más rápido cuando están más cerca del Sol y más lentamente cuando están lejos del Sol, precisamente de esta manera: supongamos que se observa un planeta en dos tiempos sucesivos cualesquiera, digamos separados una semana, y que el vector radio* se dibuja hacia el planeta para cada posición observada. El arco orbital recorrido por el planeta durante una semana y los dos vectores radio limitan cierta superficie plana, el área sombreada que se muestra en la figura 7-2. Si se hacen dos observaciones similares con una semana de separación, en una parte de la órbita más lejos del Sol (donde el planeta se mueve más lentamente), el área limitada en forma similar es exactamente la misma que en el primer caso. Así, de acuerdo con la segunda ley, la velocidad orbital de cada planeta es tal que el radio "barre" áreas iguales en tiempos iguales.

Finalmente, Kepler descubrió mucho más tarde una tercera ley; esta ley es de una categoría diferente de las otras dos, ya que trata no sólo con un planeta, sino que relaciona un planeta con otro. Esta ley dice que cuando se comparan el periodo orbital y el tamaño de la órbita de dos planetas cualesquiera, los periodos son proporcionales a la potencia 3/2 del tamaño de la órbita. En esta afirmación el periodo es el intervalo de tiempo que le lleva a un planeta completar su órbita y el tamaño se mide por la longitud del diámetro mayor de la órbita elíptica, conocido técnicamente como el eje mayor. Más sencillamente, si los planetas se movieran en círculos, como aproximadamente lo hacen, el tiempo requerido para moverse alrededor del círculo sería proporcional a la potencia 3/2 del diámetro (o el radio). Por lo tanto, las tres leyes de Kepler son:

I. Cada planeta se mueve alrededor del Sol en una elipse, con el Sol en uno de los focos.

II. El radio vector desde el Sol al planeta barre áreas iguales en intervalos iguales de tiempo.

III. Los cuadrados de los periodos de dos planetas cualesquiera son proporcionales a los cubos de los semiejes mayores de sus respectivas órbitas: $T \propto a^{3/2}$.

7-3 Desarrollo de la dinámica

Mientras Kepler descubría estas leyes, Galileo estaba estudiando las leyes del movimiento. El problema era, ¿qué hace que los planetas giren? (En esos días, una de las teorías propuestas era que los planetas giraban porque detrás de ellos iban ángeles invisibles, batiendo sus alas e impulsando los planetas hacia adelante. ¡Ustedes verán que esta teoría está ahora modificada! Resulta que para mantener los planetas girando, los ángeles invisibles deben volar en una dirección diferente y que no tienen alas. Por lo demás, es una teoría bastante similar.) Galileo descubrió un hecho muy notable acerca del movimiento, que fue esencial en la comprensión de estas leyes. Éste es el principio de *inercia*: si algo se mueve, sin que nada lo toque y sin perturbación alguna, se moverá eternamente, siguiendo a velocidad uniforme una línea recta. (*¿Por qué* se sigue moviendo? No lo sabemos, pero es así.)

---

* Un vector radio es una línea dibujada desde el Sol a cualquier punto de la órbita de un planeta.

Newton modificó esta idea, diciendo que el único modo de cambiar el movimiento de un cuerpo es usar una *fuerza*. Si un cuerpo aumenta su velocidad, una fuerza ha sido aplicada *en la dirección del movimiento*. Por otra parte, si su movimiento se cambia a una nueva *dirección*, una fuerza ha sido aplicada *lateralmente*. Así Newton agregó la idea de que se necesita una fuerza para cambiar la velocidad *o la dirección* del movimiento de un cuerpo. Por ejemplo, si una piedra está amarrada a una cuerda y está girando en un círculo, se necesita una fuerza para mantenerla en el círculo. Tenemos que *tirar* de la cuerda. De hecho, la ley es que la aceleración producida por la fuerza es inversamente proporcional a la masa o que la fuerza es proporcional a la masa por la aceleración. Mientras más masiva es una cosa, mayor es la fuerza necesaria para producir una aceleración dada. (Las masas pueden medirse colocando otras piedras al extremo de la misma cuerda y haciéndolas girar en el mismo círculo y a la misma velocidad. De este modo se encuentra que se requiere una fuerza mayor o menor, requiriendo más fuerza los objetos más masivos.)

La brillante idea que resulta de estas consideraciones es que no se necesita fuerza *tangencial* para mantener un planeta en su órbita (los ángeles no tienen que volar tangencialmente), porque el planeta seguiría en esa dirección de todos modos. Si no hubiera nada que lo perturbara, el planeta se iría en *línea recta*. Pero el movimiento se desvía de la línea en que se habría movido el cuerpo si no hubiera fuerza, siendo la desviación esencialmente en *ángulos rectos* al movimiento, no en dirección del movimiento. En otras palabras, debido al principio de inercia, la fuerza necesaria para controlar el movimiento de un planeta *alrededor* del Sol no es la fuerza alrededor del Sol sino *hacia* el Sol. (¡Si hay una fuerza hacia el Sol, éste podría ser el ángel, por supuesto!)

7-4 Ley de la gravitación de Newton

A partir de su mejor comprensión de la teoría del movimiento, Newton estimó que *el Sol* podría ser el asiento o el organismo de las fuerzas que gobiernan el movimiento de los planetas. Newton probó para sí mismo (tal vez nosotros seamos pronto capaces de probarlo) que el hecho mismo que áreas iguales sean barridas en tiempos iguales es una indicación precisa de la proposición de que todas las desviaciones son justamente *radiales* —que la ley de las áreas es una consecuencia directa de la idea de que todas las fuerzas están exactamente dirigidas *hacia el Sol*—.

A continuación, al analizar la tercera ley de Kepler, es posible demostrar que mientras más lejos esté el planeta, más débiles son las fuerzas. Si se comparan dos planetas a diferentes distancias del Sol, el análisis muestra que las fuerzas son inversamente proporcionales a los cuadrados de sus respectivas distancias. Con la combinación de las dos leyes, Newton concluyó que debe haber una fuerza, inversa al cuadrado de la distancia, dirigida en una línea entre los dos objetos.

Siendo un hombre de considerable sentido para las generalizaciones, Newton supuso, por supuesto, que esta relación se aplica más generalmente que sólo al Sol sujetando los planetas. Ya se sabía, por ejemplo, que el planeta Júpiter tenía lunas girando en torno a él, como la Luna de la Tierra gira en torno al planeta, y Newton se sintió seguro de que cada planeta sostiene sus lunas mediante una fuerza. Él conocía ya la fuerza que *nos* mantiene sobre la Tierra, de modo que propuso que era una *fuerza universal —que cada cosa atrae a las demás—*.

El problema siguiente fue si la atracción de la Tierra sobre las personas era la "misma" que sobre la Luna, es decir, inversa al cuadrado de la distancia. Si un objeto sobre la superficie de la Tierra cae 5 metros en el primer segundo después que se suelte, ¿qué distancia caerá la Luna durante el mismo tiempo? Podríamos decir que la Luna no cae en absoluto. Pero si no hubiera fuerza sobre la Luna, se escaparía en línea recta, mientras que en cambio se mueve en un círculo, de modo que realmente *cae* desde donde habría estado si no hubiera habido fuerza alguna. Podemos calcular a partir del radio de la órbita de la Luna (que es aproximadamente 384 000 km) y de cuánto tarda en ir alrededor de la Tierra (aproximadamente 29 días), cuánto se mueve la Luna sobre su órbita en un

segundo y podemos calcular entonces cuánto cae en un segundo.* Esta distancia resulta ser aproximadamente 1.3 mm en un segundo. Esto se ajusta muy bien con la ley de la inversa del cuadrado, porque el radio de la Tierra es 6 400 kilómetros, y si algo que está a 6 400 kilómetros del centro de la Tierra cae cinco metros en un segundo, algo a 384 000 kilómetros, o 60 veces más lejos, caería 1/3 600 de cinco metros, lo que también es aproximadamente 1.3 mm. Deseando poner a prueba esta teoría de la gravitación mediante cálculos similares, Newton hizo sus cálculos muy cuidadosamente y encontró una discrepancia tan grande que consideró la teoría en contradicción con los hechos y no publicó sus resultados. Seis años después las nuevas medidas del tamaño de la Tierra mostraron que los astrónomos habían estado usando una distancia a la Luna incorrecta. Cuando Newton oyó acerca de esto, hizo sus cálculos de nuevo, con las cifras correctas y obtuvo una hermosa concordancia.

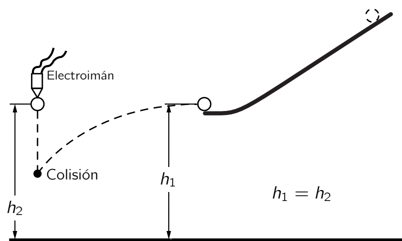


Figura 7-3. Aparato para demostrar la independencia de movimientos verticales y horizontales.

Esta idea de que la Luna "cae" es algo confusa, porque, como ven, no se *acerca* en absoluto. La idea es lo suficientemente interesante para merecer más explicación: la Luna cae en el sentido que *cae desde la línea que habría seguido si no hubiera fuerzas.* Consideremos un ejemplo en la superficie de la Tierra. Un objeto que se suelta cerca de la superficie de la Tierra caerá cinco metros en el primer segundo. Un objeto lanzado *horizontalmente* también caerá cinco metros; aun cuando se esté moviendo horizontalmente todavía caerá los cinco metros en el mismo tiempo. La figura 7-3 ilustra un aparato que demuestra esto. En la pista horizontal hay una bola que será impulsada hacia adelante en una pequeña distancia. A la misma altura hay una bola que va a caer verticalmente y hay un interruptor eléctrico arreglado de modo que en el momento que la primera bola deja la pista, se suelta la segunda bola. Que ellas llegan a la misma profundidad en el mismo tiempo está atestiguado por el hecho de que chocan en medio del aire. Un objeto como una bala, lanzado horizontalmente, puede moverse un camino largo en un segundo —tal vez 700 metros— pero siempre caerá cinco metros si es apuntado horizontalmente. ¿Qué ocurre si lanzamos una bala más y más rápido? No olviden que la superficie de la Tierra es curva. Si la disparamos lo suficientemente rápido, entonces cuando caiga cinco metros puede estar a la misma altura sobre la Tierra que lo que estuvo antes. ¿Cómo puede ser esto? Siempre cae, pero la Tierra se encorva, así que cae "alrededor" de la Tierra. El problema es: ¿qué distancia tiene que moverse en un segundo para que la Tierra esté cinco metros bajo el horizonte? En la figura 7-4 vemos la Tierra con su radio de 6 400 kilómetros y la trayectoria tangencial rectilínea que la bala tomaría si no hubiera fuerzas. Ahora, si usamos uno de esos maravillosos teoremas de la geometría, que dice que nuestra tangente es la media geométrica de las dos partes del diámetro cortado por una cuerda igual, vemos que la distancia horizontal viajada es la media geométrica de los cinco metros caídos y los 12 800 kilómetros de diámetro de la Tierra. La raíz cuadrada de $(5/1\,000) \times 12\,800$ resulta muy cercana a ocho kilómetros. Vemos así que si la bala se mueve a ocho kilómetros por segundo, continuará cayendo hacia la Tierra en la misma razón de cinco metros cada segundo, pero nunca logrará acercarse, porque la Tierra al encorvarse se aleja. Así fue como el señor Gagarin se mantuvo en el espacio viajando 40 000 kilómetros alrededor de la Tierra a ocho kilómetros por segundo aproximadamente. (Él demoró un poco más, porque estaba un poco más alto.)

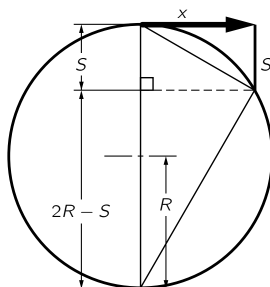


Figura 7-4. Aceleración hacia el centro de una trayectoria circular. De la geometría plana, $x/S = 2R - S/x \approx 2R/x$, donde R es el radio de la Tierra, 6 400 km; x es la distancia "recorrida horizontalmente" en un segundo; y S es la distancia "caída" en un segundo (5 metros).

Cualquier gran descubrimiento de una nueva ley es útil sólo si podemos sacar más de él que lo que ponemos. Pues bien, Newton *usó* la segunda y la tercera de las leyes de Kepler para deducir su ley de la gravitación. ¿Qué *predijo*? Primero, su análisis del movimiento de la Luna fue una predicción, porque relacionaba la caída de los objetos sobre la superficie de la Tierra con la caída de la Luna. Segundo, la pregunta es: *¿es la órbita una elipse?* Veremos en un capítulo posterior cómo es posible calcular exactamente el movimiento y uno puede probar que debe ser una elipse,† de modo que no se necesitan hechos adicionales para explicar la *primera* ley de Kepler. Así, Newton hizo su primera predicción.

* Es decir, en cuánto cae el círculo de la órbita lunar por debajo de la línea recta tangente a ésta en el punto en que estaba la Luna un segundo antes.

† La demostración no está dada en este curso.

La ley de la gravitación explica muchos fenómenos no comprendidos anteriormente. Por ejemplo, la atracción de la Luna sobre la Tierra causa las mareas, hasta entonces misteriosas. La Luna atrae al agua que está por debajo de ella y produce las mareas —la gente había pensado en eso antes, pero no fueron tan inteligentes como Newton y así pensaron que debiera haber sólo una marea durante el día—. El razonamiento era que la Luna atrae al agua por debajo de ella, produciendo una marea alta y una marea baja y, como la Tierra está rotando debajo, esto hace que la marea en un lugar suba y baje cada 24 horas. Realmente la marea sube y baja en 12 horas. Otra escuela de pensamiento afirmaba que la marea alta debería estar en el otro lado de la Tierra porque, según ellos, ¡la Luna tiraba la Tierra fuera del agua! Ambas teorías son erróneas. Realmente ocurre de este modo: la atracción de la Luna sobre la Tierra y sobre el agua está "equilibrada" en el centro. Pero el agua que está *más* cerca de la Luna es atraída más que el promedio y el agua que está más lejos de ella es atraída *menos* que el promedio. Más aún, el agua puede fluir mientras que la Tierra, que es más rígida, no puede. La verdadera descripción es una combinación de estas dos cosas.

¿Qué queremos decir por "equilibrada"? ¿Qué equilibra? Si la Luna atrae toda la Tierra hacia ella, ¿por qué la Tierra no se precipita "hacia" la Luna? Porque la Tierra hace el mismo truco que la Luna, se mueve en un círculo alrededor de un punto que está dentro de la Tierra, pero no en su centro. La Luna no gira precisamente alrededor de la Tierra; tanto la Tierra como la Luna giran en torno a una posición central, cayendo ambas hacia esta posición común, como se muestra en la figura 7-5. Este movimiento alrededor de un centro común es lo que equilibra la caída de cada una. De modo que tampoco la Tierra se mueve en una línea recta; viaja en un círculo. El agua en la parte más alejada está "desequilibrada", porque la atracción de la Luna allí es más débil que en el centro de la Tierra, donde equilibra justamente la "fuerza centífruga". El resultado de este desequilibrio es que el agua sube alejándose del centro de la Tierra. En el lado cercano, la atracción de la Luna es más fuerte y el desequilibrio está en dirección opuesta en el espacio, pero de nuevo *alejándose* del centro de la Tierra. El resultado neto es que tenemos *dos* subidas de marea.

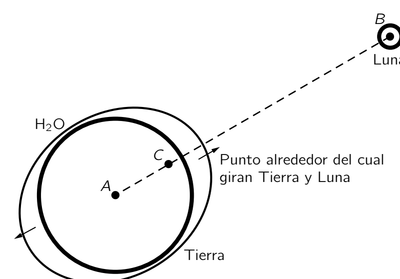


Figura 7-5. El sistema Tierra-Luna, con mareas.

7-5 Gravitación universal

¿Qué más podemos comprender al comprender la gravedad? Todos saben que la Tierra es redonda. ¿Por qué es redonda la Tierra? Esto es fácil: debido a la gravitación. ¡Puede comprenderse que la Tierra sea redonda simplemente porque cada cosa atrae a cada cosa y así se ha atraído juntándose a sí misma lo más que ha podido! Si vamos aún más lejos, la Tierra no es *exactamente* una esfera porque está rotando y esto introduce efectos centrífugos que tienden a oponerse a la gravedad cerca del ecuador. Resulta que la Tierra debería ser elíptica, y nosotros incluso obtenemos la forma correcta para la elipse. Podemos así deducir que el Sol, la Luna y la Tierra deberían ser (aproximadamente) esferas, justamente a partir de la ley de la gravitación.

¿Qué más pueden hacer con la ley de la gravitación? Si miramos las lunas de Júpiter podemos entender todo acerca del modo en que se mueven alrededor del planeta. A propósito, hubo una vez cierta dificultad con las lunas de Júpiter que es digno hacer notar. Estos satélites fueron estudiados con mucho cuidado por Rømer, quien notó que a veces las lunas parecían estar adelantadas respecto de su horario, y a veces atrasadas. (Se pueden encontrar sus horarios esperando un tiempo muy largo y encontrando lo que demoran en promedio las lunas en girar.) Pues bien, ellas se *adelantaban* cuando Júpiter estaba particularmente *cerca* de la Tierra y se *atrasaban* cuando Júpiter estaba más *lejos* de la Tierra. Esto habría sido algo muy difícil de explicar con la teoría de la gravitación —habría sido, de hecho, la muerte de esta maravillosa teoría, si no hubiera otra explicación—. Si una ley no funciona siquiera en *un lugar* donde debiera hacerlo, está simplemente equivocada. Pero la razón de esta discrepancia era muy simple y hermosa: requiere un pequeño instante *ver* las lunas de Júpiter debido al tiempo que demora la luz en viajar de Júpiter a la Tierra. Cuando Júpiter está

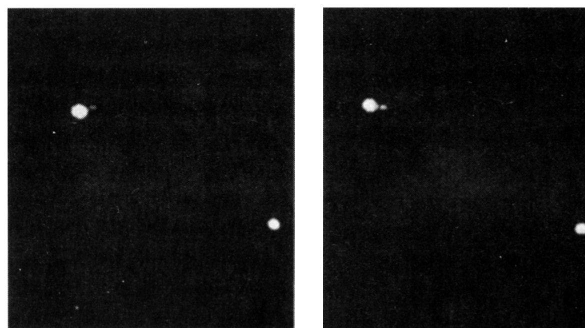


Figura 7-6. Un sistema de estrella doble

más cerca de la Tierra, el tiempo es un poco menor, y cuando está más lejos de la Tierra, el tiempo es mayor. Ésta es la razón por la que las lunas parecen estar, en promedio, un poco adelantadas o un poco atrasadas, según si están más cerca o más lejos de la Tierra. Este fenómeno demostró que la luz no viaja instantáneamente, y proporcionó el primer cálculo de la velocidad de la luz. Esto fue hecho en 1676.

Si todos los planetas se empujan y tiran entre sí, la fuerza que controla, digamos, Júpiter al ir alrededor del Sol, no es precisamente la fuerza desde el Sol; también hay un tirón desde, digamos, Saturno. Esta fuerza realmente no es grande, ya que el Sol es mucho más masivo que Saturno, pero hay *cierta* atracción, de modo que la órbita de Júpiter no debía ser una elipse perfecta y no lo es; está ligeramente corrida y "oscila" alrededor de una órbita elíptica correcta. Tal movimiento es un poco más complicado. Se hicieron intentos de analizar los movimientos de Júpiter, Saturno y Urano sobre la base de la ley de la gravitación. Se calcularon los efectos de cada uno de estos planetas sobre los demás, para ver si las leves desviaciones e irregularidades de estos movimientos podrían ser completamente comprendidas a partir de esta única ley. Para Júpiter y Saturno todo andaba bien, pero Urano estaba "raro". Se comportaba de un modo peculiar. No se movía en una elipse exacta, pero eso era comprensible debido a las atracciones de Júpiter y Saturno. Pero aun haciendo concesiones sobre estas atracciones, Urano *todavía* no marchaba bien, de modo que las leyes de la gravitación estaban en peligro de zozobrar, una posibilidad que no podía ser descartada. Dos hombres, Adams y Le Verrier, en Inglaterra y Francia, independientemente llegaron a otra posibilidad: tal vez haya *otro* planeta, oscuro e invisible, que los hombres no habían visto. Este planeta, N,

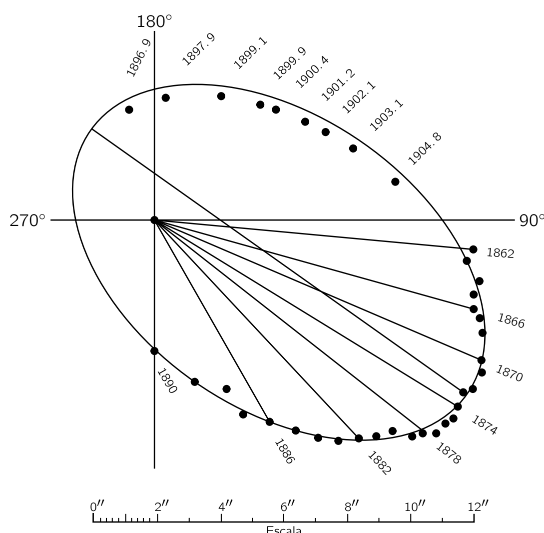


Figura 7-7. Órbita de Sirio B respecto a Sirio A.

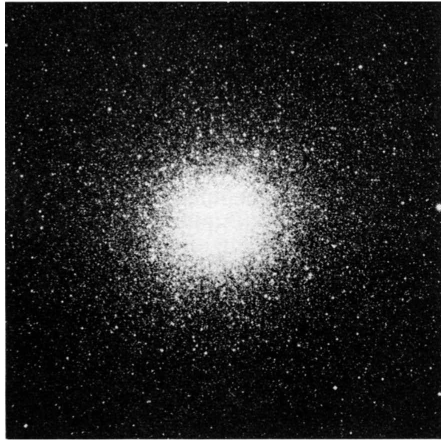


Figura 7-8. Un cúmulo globular de estrellas.

atraería a Urano. Calcularon dónde debería estar tal planeta para causar las perturbaciones observadas. Enviaron mensajes a los respectivos observatorios, diciendo: "Señores, dirijan sus telescopios a tal y tal lugar y verán un nuevo planeta". A menudo depende de con quién estén ustedes trabajando para que les presten atención o no. Ellos le pusieron atención a Le Verrier; ¡miraron y ahí estaba el planeta N! Entonces otros observatorios también se apresuraron a mirar en los días siguientes y también lo vieron.

Este descubrimiento mostró que las leyes de Newton son absolutamente correctas en el sistema solar; pero ¿pueden extenderse más allá de las distancias relativamente pequeñas de los planetas más cercanos? La primera prueba está en esta pregunta: ¿se atraen las *estrellas entre sí* al igual que los planetas? Tenemos una definitiva evidencia que lo hacen en las *estrellas dobles.* La figura 7-6 muestra una estrella doble —dos estrellas muy juntas— (hay además una tercera estrella en el cuadro de modo que sabemos que la fotografía no ha sido girada). Además, las estrellas se muestran tal como aparecieron varios años más tarde. Vemos que en relación con la estrella "fija" el eje del par ha rotado, es decir, las dos estrellas han ido una en torno a la otra. ¿Rotarán de acuerdo a las leyes de Newton? Cuidadosas medidas de las posiciones relativas de un tal sistema de estrellas dobles se muestran en la figura 7-7. Vemos allí una bella elipse, las medidas parten en 1862 y dando la vuelta completa hasta 1904 (ahora debe haber completado una vuelta más). Todo coincide con las leyes de Newton excepto que la estrella Sirio A *no está en el foco.* ¿Por qué será? Porque el plano de la elipse no está en el "plano del cielo". No estamos mirando en ángulo recto



Figura 7-9. Una galaxia.

al plano de la órbita, y cuando una elipse se ve inclinada, sigue siendo una elipse, pero el foco ya no está más en el mismo lugar. Así podemos analizar las estrellas dobles, moviéndose una respecto a la otra, de acuerdo a las exigencias de la ley gravitacional.

Que la ley de la gravitación es válida para distancias aún mayores se indica en la figura 7-8. Si alguien no puede ver la gravitación actuando aquí es que no tiene alma. Esta figura muestra una de las más bellas cosas en el cielo: un cúmulo globular de estrellas. Todos los puntos son estrellas. Aunque parecen estar agrupadas en forma compacta hacia el centro, esto se debe a la falibilidad de nuestros instrumentos. En realidad, las distancias entre las estrellas, aun las más centrales, son muy grandes y raramente chocan. Hay muchas más estrellas en el interior que alejadas, y a medida que nos alejamos hay cada vez menos. Es obvio que hay una atracción entre estas estrellas. Es evidente que la gravitación existe a estas enormes dimensiones, tal vez 100 000 veces el tamaño del sistema solar. Vamos ahora más lejos y observemos una *galaxia entera*, mostrada en la figura 7-9. La forma de esta galaxia indica una tendencia evidente de la materia a aglomerarse. Por supuesto, no podemos probar aquí que la ley sea precisamente la de la inversa del cuadrado, sólo que aún existe una atracción a esta enorme dimensión que mantiene todo junto. Alguien podría decir: “Bien, todo esto es muy ingenioso, pero, ¿por qué no es justamente una bola?” Porque está *girando* y tiene *momentum angular* que no debe ceder al contraerse; debe contraerse principalmente en un plano. (A propósito, si andan buscando un buen problema,



Figura 7-10. Un cúmulo de galaxias.



Figura 7-11. Una nube de polvo interestelar.



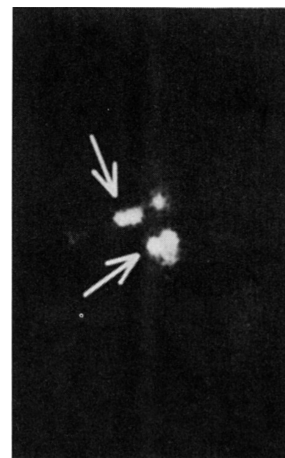


Figura 7-12. ¿La formación de nuevas estrellas?

los detalles exactos de cómo se forman los brazos y qué determina la forma de estas galaxias no se han elaborado aún.) Sin embargo, es claro que la forma de la galaxia se debe a la gravitación, aunque las complejidades de su estructura no nos han permitido aún analizarlo completamente. En una galaxia tenemos una escala de tal vez 50 000 a 100 000 años luz. La distancia de la Tierra al Sol es $8\frac{1}{3}$ *minutos* luz, de modo que pueden ver lo grandes que son estas dimensiones.

La gravedad parece existir aún a dimensiones mayores como se indica en la figura 7-10, que muestra muchas cosas "pequeñas" aglomeradas. Éste es un *cúmulo de galaxias*, tal como un cúmulo de estrellas. Así las galaxias se atraen unas a otras a tales distancias que también están amontonadas en cúmulos. Tal vez la gravitación exista aun a distancias de *decenas de millones* de años luz; por lo que sabemos ahora, la gravedad parece extenderse siempre inversamente con el cuadrado de la distancia.

No sólo logramos entender las nebulosas, sino que a partir de la ley de gravitación podemos obtener además algunos planteamientos acerca del origen de las estrellas. Si tenemos una gran nube de polvo y gas, como en la figura 7-11, las atracciones gravitacionales de las partes de polvo entre sí pueden hacerlas formar pequeños bultos. Escasamente visibles en la figura hay unos "pequeños" puntos negros que pueden ser el comienzo de la acumulación de polvo y gases, que debido a su gravitación empiezan a formar estrellas. Si hemos visto formarse alguna vez una estrella o no, es aún discutible. La figura 7-12 muestra alguna evidencia que sugiere que sí. A la izquierda hay una fotografía tomada en 1947 de una región de gas con algunas estrellas, y a la derecha hay otra fotografía, tomada sólo siete años después, que muestra dos nuevas manchas brillantes. ¿Se ha acumulado gas, ha actuado la gravedad lo suficientemente fuerte como para juntarlo en una bola lo suficientemente grande para hacer que la reacción nuclear estelar comience en el interior y la transforme en una estrella? Quizá sí y quizá no. No es razonable que en sólo siete años hayamos sido tan afortunados como para ver convertirse una estrella en una forma visible; ¡es mucho menos probable que viéramos *dos*!

7-6 El experimento de Cavendish

La gravitación, por lo tanto, se extiende sobre enormes distancias. Pero si hay una fuerza entre *cualquier* par de objetos, deberíamos poder medir la fuerza entre nuestros propios objetos. En vez de tener que mirar las estrellas moverse una en torno de la otra, ¿por qué no tomar una bola de plomo y una de marfil y observar que la de marfil va hacia la de plomo? La dificultad de este experimento cuando se hace de una manera tan simple es la gran debilidad o delicadeza de la fuerza. Debe hacerse con extremo cuidado, lo que significa cubrir el aparato para mantener fuera el aire, asegurarse de que no esté cargado eléctricamente, etc.; entonces la fuerza puede medirse. El primero en medirla fue Cavendish con un aparato que está indicado esquemáticamente en la figura 7-13. Éste demostró primero la fuerza directa entre dos grandes bolas fijas de plomo y dos bolas más pequeñas de plomo en los extremos de un brazo suspendido de una fibra muy fina, llamada fibra de torsión. Midiendo cuánto se tuerce la fibra se puede medir la magnitud de la fuerza, verificar que es inversamente proporcional al cuadrado de la distancia y determinar su intensidad. Así se puede determinar precisamente el coeficiente G de la fórmula

$$F = G\,\frac{mm'}{r^2}.$$

Todas las masas y distancias se conocen. Ustedes dirán "Ya lo sabíamos para la Tierra". Sí, pero no conocíamos la *masa* de la Tierra. ¡Conociendo G a partir de este experimento y conociendo la fuerza con que la Tierra atrae, podemos averiguar indirectamente cuál es el valor de la masa de la Tierra! Este experimento se ha llamado "pesar la Tierra" por algunas personas, y puede ser usado para

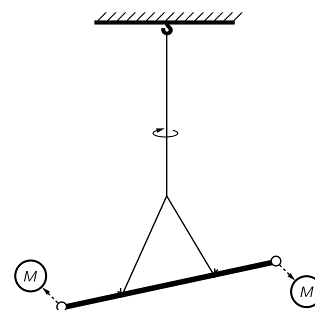


Figura 7-13. Un diagrama simplificado del aparato usado por Cavendish para verificar la ley de gravitación universal para objetos pequeños y medir la constante gravitacional G.

determinar el coeficiente G de la ley de la gravedad. Éste es el único modo en que se puede determinar la masa de la Tierra. G resultó ser

$$6.670 \times 10^{-11} \text{ newton} \cdot \text{m}^2/\text{kg}^2.$$

Es difícil exagerar la importancia del efecto producido en la historia de la ciencia por este gran éxito de la teoría de la gravitación. ¡Comparen la confusión, la falta de confianza, el conocimiento incompleto que prevalecía en los primeros tiempos, cuando había interminables debates y paradojas, con la claridad y simplicidad de esta ley; este hecho que todas las lunas y planetas y estrellas tengan una *regla tan simple* que los gobierna, y más aún que el hombre pueda *entenderla* y deducir cómo se mueven los planetas! Ésta es la razón del éxito de las ciencias en los años siguientes, porque dio la esperanza de que los demás fenómenos del mundo también tengan leyes tan bellamente simples.

7-7 ¿Qué es la gravedad?

¿Pero es ésta una ley tan simple? ¿Cuál es su mecanismo? Todo lo que hemos hecho es describir *cómo* se mueve la Tierra alrededor del Sol, pero no se dijo *qué la hace mover.* Newton no hizo hipótesis sobre esto; él quedó satisfecho con encontrar *lo que* hacía sin entrar en su mecánica. *Nadie ha dado desde entonces ningún mecanismo.* Es característico de las leyes físicas que tengan este carácter abstracto. La ley de conservación de la energía es un teorema concerniente a cantidades que deben calcularse y sumarse, sin mención del mecanismo, y en forma parecida las grandes leyes de la mecánica son leyes matemáticas cuantitativas, para las cuales no hay mecanismo disponible. ¿Por qué podemos usar las matemáticas para describir la naturaleza sin un mecanismo tras ella? Nadie lo sabe. Tenemos que proseguir porque de esa manera descubrimos más.

Se han sugerido muchos mecanismos para la gravitación. Es interesante considerar uno de éstos, ya que mucha gente ha pensado en ello de tiempo en tiempo. Al comienzo, uno se excita y se pone contento cuando lo "descubre", pero pronto encuentra que no es correcto. Fue descubierto alrededor de 1750. Supongamos que hay muchas partículas moviéndose en el espacio a una velocidad muy grande y en todas direcciones y que son sólo ligeramente absorbidas al moverse a través de la materia. Cuando *son* absorbidas, dan un impulso a la Tierra. Sin embargo, ya que hay tantas moviéndose de una forma y otra, los impulsos se compensan. Pero cuando el Sol está cerca, las partículas que vienen hacia la Tierra a través del Sol son parcialmente absorbidas, de modo que hay menos partículas viniendo del Sol que viniendo del otro lado. Por lo tanto, la Tierra siente un impulso neto hacia el Sol y no 1e lleva mucho a uno ver que es inversamente proporcional al cuadrado de las distancias —debido a la variación del ángulo sólido que el Sol subtiende cuando variamos la distancia—. ¿Qué anda mal en este mecanismo? Comprende algunas nuevas consecuencias que *no son verdaderas.* Esta idea particular tiene la siguiente dificultad: la Tierra al moverse alrededor del Sol chocaría con más partículas que vienen de su lado delantero que de su lado trasero (¡cuando corren en la lluvia, la lluvia en su cara es más fuerte que en la parte de atrás de su cabeza!). Por lo tanto, se habría dado más impulso a la Tierra por delante y la Tierra sentiría una *resistencia al movimiento* y se estaría retardando en su órbita. Uno puede calcular cuánto tomaría para que la Tierra se detenga como resultado de esta resistencia y este tiempo no sería lo suficientemente largo como para que la Tierra permanezca en su órbita, de modo que este mecanismo no funciona. No se ha inventado ningún mecanismo que "explique" la gravedad sin que prediga algunos otros fenómenos que *no* existen.

A continuación discutiremos la posible relación de la gravitación con otras fuerzas. No existe explicación de la gravitación en términos de otras fuerzas en el presente. No es un aspecto de la electricidad ni nada como eso, de modo que no tenemos explicación. Sin embargo, la gravitación y otras fuerzas son muy similares, y es interesante notar analogías. Por ejemplo, la fuerza eléctrica entre dos objetos cargados se parece exactamente a la ley de la gravitación. La fuerza

eléctrica es una constante con un signo menos, por el producto de las cargas, y varía inversamente con el cuadrado de la distancia. Está en dirección opuesta —iguales se repelen—. ¿Pero no es sumamente notable que las dos leyes contienen la misma función de la distancia? Tal vez la gravitación y la electricidad están mucho más íntimamente relacionadas de lo que pensamos. Se han hecho muchos intentos de unificarlas; la así llamada teoría del campo unificado es sólo un intento muy elegante de combinar electricidad y gravitación; pero al comparar la gravitación y la electricidad, lo más interesante es la *intensidad relativa* de las fuerzas. Cualquier teoría que contenga ambos debe deducir además la intensidad de la gravedad.

Si tomamos, en ciertas unidades naturales, la repulsión de dos electrones (carga universal de la naturaleza) debida a la electricidad, y la atracción de dos electrones debida a sus masas, podemos medir el cociente entre la repulsión eléctrica y la atracción gravitacional. El cociente es independiente de la distancia y es una constante fundamental de la naturaleza. La figura 7-14 muestra ese cociente. ¡La atracción gravitacional comparada con la repulsión eléctrica entre dos electrones es 1 dividido entre 4.17×10^{42}! La pregunta es, ¿de dónde salió ese número tan grande? No es accidental, como el cociente entre el volumen de la Tierra y el volumen de una pulga. Hemos considerado dos aspectos naturales de una misma cosa, un electrón. Este número fantástico es una constante natural, de modo que encierra algo de naturaleza profunda. ¿De dónde podría salir un número tan tremendo? Algunos dicen que algún día encontraremos la "ecuación universal", y en ella una de las raíces será este número. Es muy difícil encontrar una ecuación para la cual un número tan fantástico sea una raíz natural. Se han pensado otras posibilidades; una es relacionarla con la edad del universo. Evidentemente, debemos encontrar *otro* gran número en alguna parte. Pero ¿queremos indicar la edad del universo en *años*? No, porque los años no son "naturales"; fueron ideados por los hombres. Como ejemplo de algo natural, consideremos el tiempo que demora la luz en atravesar un protón, 10^{-24} segundos. Si comparamos este tiempo con la *edad del universo*, 2×10^{10} años, la respuesta es 10^{-42}. Tiene alrededor del mismo número de ceros, de modo que se ha propuesto que la constante gravitacional está relacionada con la edad del universo. Si tal fuera el caso, la constante gravitacional debería cambiar con el tiempo, porque a medida que el universo se hace más viejo, la razón entre la edad del universo y el tiempo que demora la luz en cruzar un protón irá aumentando gradualmente. ¿Es posible que la constante gravitacional *esté* cambiando con el tiempo? Por supuesto que los cambios serían tan pequeños que es bien difícil estar seguro.

Una prueba que podemos idear es determinar cuál habría sido el efecto del cambio durante los últimos 10^9 años, que es aproximadamente la edad de la primera vida sobre la Tierra y un décimo de la edad del universo. En este tiempo, la constante gravitatoria habría aumentado cerca de un 10 %. Resulta que si consideramos la estructura del Sol —el equilibrio entre el peso de su materia y la rapidez a la cual la energía radiante se genera dentro de él— podemos deducir que si la gravedad fuera 10 % mayor, el Sol debería ser mucho más que 10 % más brillante —¡en la *sexta potencia* de la constante de gravedad!—. Si calculamos lo que le sucede a la órbita de la Tierra cuando la gravedad cambia, encontramos que entonces la Tierra estaba *más cerca*. En total, la Tierra estaría cerca de 100 grados centígrados más caliente y toda el agua no habría estado en el mar, sino como vapor en el aire, de modo que la vida no habría comenzado en el mar. Así, *no* creemos que la constante de gravedad esté cambiando con la edad del universo. Pero tales argumentos como el que acabamos de dar no son muy convincentes y el asunto no está completamente terminado.

Es un hecho que la fuerza de gravitación es proporcional a la *masa*, la cantidad que es fundamentalmente una medida de la *inercia* —de cuánto cuesta sostener algo que está girando en un círculo—. Por lo tanto, dos objetos, uno pesado y otro liviano, girando alrededor de un objeto mayor en el mismo círculo a la misma velocidad, debido a la gravedad, permanecerán juntos porque ir en un círculo *requiere* una fuerza que es más intensa para una masa más grande. Esto

$$\frac{\text{Gravitation Attraction}}{\text{Electrical Repulsion}} = 1/4.17 \times 10^{42}$$

Figura 7-14. Las intensidades relativas de las interacciones eléctricas y gravitacional entre dos electrones.

es, la gravedad es mayor para una masa dada en *precisamente la proporción justa*, de modo que los dos objetos giran juntos. Si un objeto estuviera dentro de otro, *permanecería* adentro; es un equilibrio perpetuo. Por lo tanto, Gagarín o Titov encontrarían las cosas "sin peso" dentro de una nave espacial; si soltaran un pedazo de tiza, por ejemplo, éste giraría alrededor de la Tierra exactamente en la misma forma que toda la nave espacial, y parecería que permanece suspendido en el espacio delante de ellos. Es muy interesante que esta fuerza es *exactamente* proporcional a la masa con gran precisión, porque si no fuera cabalmente proporcional habría algún efecto por el cual la inercia y el peso diferirían. La ausencia de tal efecto se ha comprobado con gran exactitud con un experimento hecho primero por Eötvös en 1909 y más recientemente por Dicke. Para todas las sustancias consideradas, las masas y pesos eran exactamente proporcionales en una parte en 1 000 millones, o menos. Éste es un experimento notable.

7-8 Gravedad y relatividad

Otro tópico que merece discusión es la modificación de Einstein a la ley de gravitación de Newton. ¡A pesar de todo el entusiasmo que produjo, la ley de gravitación de Newton no es correcta! Fue modificada por Einstein para tomar en cuenta la teoría de la relatividad. De acuerdo con Newton, el efecto gravitacional es instantáneo, esto es, si moviéramos una masa, sentiríamos al instante una nueva fuerza debida a la nueva posición de la masa; por este medio podríamos enviar señales con velocidad infinita. Einstein anticipó argumentos que sugieren que *no podemos enviar señales más rápidas que la velocidad de la luz*, de modo que la ley de la gravitación debe estar equivocada. Al corregirla para considerar los atrasos, tenemos una nueva ley llamada ley de gravitación de Einstein. Un aspecto de esta nueva ley, que es muy fácil de entender, es éste: en la teoría de la relatividad de Einstein, todo lo que tiene *energía* tiene masa —masa en el sentido que es atraída gravitacionalmente—. Aun la luz, que tiene una energía, tiene una "masa". Cuando un haz de luz, que contiene energía, pasa cerca del Sol, hay una atracción sobre él por el Sol. Así la luz no se mueve en línea recta, sino que es desviada. Por ejemplo, durante un eclipse de Sol, las estrellas que están rodeando al Sol aparecerían desplazadas de donde debieran estar si el Sol no estuviera ahí, y esto ha sido observado.

Finalmente, comparemos la gravitación con otras teorías. En los años recientes hemos descubierto que toda masa está hecha de pequeñas partículas y que hay muchas clases de interacciones, como fuerzas nucleares, etc. Se ha encontrado que ninguna de estas fuerzas nucleares o eléctricas explica la gravitación. Los aspectos cuánticos de la naturaleza no se han aplicado a la gravitación. Cuando la escala es tan pequeña que necesitamos de los efectos cuánticos, los efectos gravitacionales son tan débiles, que no se ha desarrollado todavía la necesidad de una teoría cuántica de la gravitación. Por otra parte, para la consistencia de nuestras teorías físicas sería importante ver si la ley de Newton modificada en la ley de Einstein puede ser modificada aún más para ser consistente con el principio de incertidumbre. Esta última modificación no ha sido aún completada.

8

Movimiento

8-1 Descripción del movimiento

A fin de encontrar las leyes que gobiernan los diversos cambios que experimentan los cuerpos a medida que el tiempo transcurre, debemos estar en condiciones de *describir* los cambios y tener alguna manera de registrarlos. El cambio más simple observable en un cuerpo es el cambio aparente en su posición respecto al tiempo, el cual llamamos movimiento. Consideremos algún objeto sólido con una marca permanente que podemos observar, que llamaremos punto. Discutiremos el movimiento del pequeño marcador, el cual podría ser la tapa del radiador de un automóvil o el centro de una pelota que está cayendo, y trataremos de describir el hecho de que se mueve y cómo se mueve.

Estos ejemplos pueden parecer triviales, pero en la descripción del cambio entran muchas sutilezas. Algunos cambios son más difíciles de describir que el movimiento de un punto en un objeto sólido, por ejemplo la velocidad de deriva de una nube que se mueve muy lentamente, pero formándose o evaporándose rápidamente, o el cambio de opinión de una mujer. No conocemos una manera simple para analizar un cambio de opinión, pero ya que la nube se puede representar o describir por muchas moléculas, en principio, tal vez podamos describir el movimiento de la nube, describiendo el movimiento de cada una de sus moléculas. De la misma manera, quizá, los cambios en la opinión pueden tener un paralelo con los cambios de los átomos dentro del cerebro, pero aún no tenemos tal conocimiento.

De todas maneras, esto es el motivo por el cual comenzamos con el movimiento de puntos; tal vez debiéramos imaginarlos como átomos, pero es probablemente mejor ser más imprecisos al comienzo y simplemente pensar en alguna especie de pequeños objetos, es decir, pequeños cuerpos comparados con la distancia recorrida. Por ejemplo, para describir el movimiento de un automóvil que va a 100 kilómetros, no tenemos que distinguir entre la parte delantera y la parte trasera del automóvil. Sin duda, hay pequeñas diferencias, pero para propósitos aproximados diremos “el automóvil” y del mismo modo no importa que nuestros puntos, no sean puntos absolutos; para lo que nos interesa no es necesario ser extremadamente preciso. También, mientras echamos una primera mirada a este tema, vamos a olvidar las tres dimensiones del mundo. Solamente nos concentraremos sobre el movimiento en una dirección, como en un automóvil sobre una carretera. Volveremos a las tres dimensiones después de ver cómo se describe el movimiento en una dimensión. Ahora bien, ustedes pueden decir: “Todo esto es algo trivial” y en verdad lo es. ¿Cómo podemos describir tal movimiento unidimensional —digamos, de un automóvil—? Nada podría ser más simple. Entre las muchas maneras posibles, una sería la siguiente. Para determinar la posición del automóvil en diferentes tiempos, medimos su distancia desde el punto de partida y anotamos todas las observaciones. En la tabla 8-1, s representa la distancia del auto en metros, desde el punto de partida, y t representa el tiempo en minutos. La primera línea de la tabla representa la distancia cero y el tiempo cero —el coche aún no ha partido—. Al minuto de su partida ha recorrido 400 metros. Después de dos minutos, va más lejos —noten que ha alcanzado mayor distancia en el segundo minuto—, ha acelerado; pero algo sucedió entre 3 y 4 y más aún cerca de los 5 —¿tal vez se detuvo en un semáforo?—. Luego su velocidad aumenta otra vez y recorre 4 300 metros al término de seis minutos, 6 000 metros al término de siete minutos, y 7 850 metros

Tabla 8-1

t (min)	s (m)
0	0
1	400
2	1 300
3	3 000
4	3 170
5	3 200
6	4 300
7	6 000
8	7 850
9	8 000

en ocho minutos; a los nueve minutos ha avanzado solamente 8 000 metros, debido a que en el último minuto fue detenido por un policía.

Ésta es una manera de describir el movimiento. Otra manera es por medio de una gráfica. Si anotamos el tiempo horizontalmente y la distancia verticalmente, obtenemos una curva parecida a la indicada en la figura 8-1. Cuando el tiempo aumenta, la distancia se incrementa, al comienzo muy lentamente y después más rápido, y otra vez muy lentamente por un pequeño lapso a los cuatro minutos; entonces aumenta otra vez durante unos pocos minutos y finalmente, a los nueve minutos, parece haber detenido su aumento. Estas observaciones pueden hacerse de la gráfica, sin una tabla. Evidentemente, para una descripción completa uno tendría que saber también dónde está el auto en las marcas de medio minuto, pero suponemos que la gráfica significa algo, que el auto tiene alguna posición en todos los instantes intermedios.

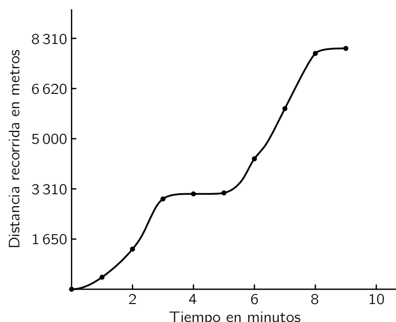


Figura 8-1. Gráfica de la distancia en función del tiempo para el auto.

El movimiento de un auto es complicado. Como otro ejemplo, usaremos algo que se mueve de una manera más simple, que siga leyes más simples: una pelota cayendo. La tabla 8-2 da el tiempo en segundos y la distancia en metros para un cuerpo que cae. A cero segundos la pelota parte desde cero metros, y al final de un segundo ha caído cinco metros. Al final de dos segundos ha caído 20 metros, al final de tres segundos 45 metros, y así sucesivamente; si se grafica los números tabulados obtenemos la bella curva parabólica indicada en la figura 8-2. La fórmula para esta curva puede ser escrita en la forma:

Tabla 8-2

t (s)	s (m)
0	0
1	5
2	20
3	45
4	80
5	125
6	180

$$s = 5t^2. \tag{8.1}$$

Esta fórmula nos permite calcular la distancia en cualquier instante. Pueden decir que también debe haber una fórmula para el primer gráfico. Realmente, uno puede escribir tal fórmula en forma abstracta,

$$s = f(t), \tag{8.2}$$

lo cual significa que s es alguna cantidad que depende de t o, en términos matemáticos, s es una función de t. Ya que nosotros no conocemos cuál es la función, no hay manera de escribirla en forma algebraica definida.

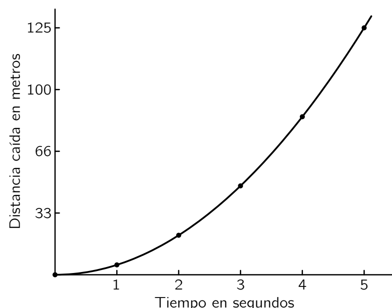


Figura 8-2. Gráfica de la distancia en función del tiempo para un cuerpo que cae.

Acabamos de ver dos ejemplos de movimiento, adecuadamente descritos con ideas muy simples, sin sutilezas. Sin embargo, *hay* sutilezas —y varias—. En primer lugar, ¿qué entendemos por *tiempo* y *espacio*? Resulta que estos profundos problemas filosóficos tienen que ser analizados muy cuidadosamente en la física, y esto no es tan fácil de hacer. La teoría de la relatividad demuestra que nuestras ideas de espacio y tiempo no son tan simples como uno pueda pensar a primera vista. Sin embargo, para lo que nos interesa, para la exactitud que necesitamos al principio, no necesitamos ser muy cuidadosos en definir exactamente las cosas. Tal vez digan: "Esto es una cosa terrible; yo aprendí que en la ciencia tenemos que definir *cada cosa* en forma precisa", ¡No podemos definir *ninguna cosa* en forma precisa! Si lo intentamos, caemos dentro de esa parálisis del pensamiento que le ocurre a los filósofos, quienes se sientan frente a frente, uno diciendo al otro: "¡Usted no sabe de lo que está hablando!" El segundo dice: "¿Qué entiende usted por *saber*? ¿Qué entiende por *hablar*? ¿Qué entiende por *usted*?", y así sucesivamente. A fin de poder conversar en forma constructiva tenemos que convenir precisamente que estamos conversando aproximadamente de la misma cosa. Ustedes saben tanto acerca del tiempo como necesitamos por ahora, pero recuerden que hay algunas sutilezas que tienen que ser discutidas, las analizaremos más tarde.

Otra sutileza ya mencionada es que sería posible imaginar que el punto móvil que estamos observando está siempre ubicado en alguna parte (por supuesto, cuando lo estamos observando, está ahí, pero puede ser que cuando miremos hacia otro lado no esté ahí). Resulta que en el movimiento de los átomos esa idea también es falsa —no podemos encontrar un marcador en un átomo y verlo moverse—. A esas sutilezas tendremos que llegar en la mecánica cuántica. Pero primero vamos a aprender cuáles son los problemas antes de introducir las complicaciones, y *entonces* estaremos en una situación mejor para hacer

correcciones a la luz del más reciente conocimiento del tema. Por lo tanto, tomaremos un punto de vista simple acerca del tiempo y del espacio. Sabemos lo que son estos conceptos de una manera aproximada, y los que han manejado un automóvil saben lo que significa velocidad.

8-2 Velocidad

Aunque sabemos aproximadamente lo que significa "velocidad", hay aún algunas sutilezas bastante profundas; tengan en cuenta que los estudiosos griegos nunca pudieron describir en forma adecuada los problemas relativos a la velocidad. La sutileza nace cuando tratamos de comprender exactamente qué se entiende por "velocidad". Los griegos se confundieron mucho con esto, y una nueva rama de la matemática tuvo que ser descubierta además de la geometría y el álgebra de los griegos, árabes y babilonios. Como ilustración de la dificultad, tratemos de resolver este problema con álgebra pura. Se está inflando un globo de modo que el volumen del globo aumenta a razón de 100 cm^3 por segundo, ¿a qué velocidad está aumentando el radio cuando el volumen es de 1 000 cm^3? Los griegos se embrollaron bastante con tales problemas, siendo ayudados, por supuesto, por algunos griegos muy confusos. Para mostrar que existían dificultades al razonar acerca de la velocidad en esa época, Zenón produjo un gran número de paradojas, de las cuales mencionaremos una para ilustrar este punto en el cual hay obvias dificultades en las ideas acerca del movimiento. "Escuchen —dice— el siguiente razonamiento: Aquiles corre 10 veces más rápido que una tortuga; sin embargo, nunca puede alcanzar a la tortuga. Para ello, supongan que inician una carrera donde la tortuga está 100 metros delante de Aquiles; entonces cuando Aquiles ha corrido los 100 metros al lugar donde estaba la tortuga, la tortuga ha avanzado 10 metros, habiendo corrido un décimo de rápido. Ahora, Aquiles tiene que correr otros 10 metros para alcanzar a la tortuga, pero al llegar al final de esa carrera, encuentra que la tortuga está aún a un metro delante de él; corriendo otro metro, encuentra a la tortuga 10 centímetros delante, y así sucesivamente, *hasta el infinito*. Por lo tanto, en cualquier instante la tortuga está siempre delante de Aquiles y Aquiles nunca puede alcanzar a la tortuga." ¿Dónde está el error en esto? Está en que una cantidad finita de *tiempo* puede ser dividida en un número infinito de partes, como una longitud de línea puede dividirse en un número infinito de partes dividiéndola repetidamente en dos. Y así, aunque hay un número infinito de pasos (en el razonamiento) hasta el punto en el cual Aquiles alcanza a la tortuga, no significa que haya una cantidad infinita de *tiempo*. Podemos ver con este ejemplo que hay en verdad algunas sutilezas en el razonamiento acerca de la velocidad.

A fin de comprender las sutilezas en una forma clara, les recordamos un chiste que seguramente deben haber oído. En el lugar donde un policía para a la dama en el auto, el policía se le acerca y dice: "¡Señora, usted iba a 100 kilómetros por hora!" Ella dice: "Eso es imposible, señor, he estado viajando sólo siete minutos. Es ridículo, ¿cómo puedo ir a 100 kilómetros por hora cuando no he viajado una hora?" ¿Cómo responderían si fueran el policía? Por supuesto, si fueran realmente el policía no habría sutilezas; es muy simple, dirían: "¡Cuénteselo al juez!" Pero supongamos que nosotros no tenemos esa salida y hacemos un ataque intelectual más honesto al problema y tratamos de explicar a esta dama lo que entendemos por la idea de que ella fuera a 100 kilómetros por hora. Precisamente, ¿qué *entendemos*? Decimos: "Lo que entendemos, señora, es esto: si usted siguiera yendo de la misma manera como iba ahora, en la hora siguiente habría recorrido 100 kilómetros". Ella podría decir: "Bien, mi pie no estaba sobre el acelerador y el auto estaba deteniéndose, así que si yo continuara yendo de esa manera no recorrería 100 kilómetros". También podemos considerar la pelota que cae y suponer que queremos conocer su velocidad en el tiempo de tres segundos, si la pelota sigue moviéndose de la manera en que lo está haciendo. ¿Qué significa eso, seguir *acelerando*, ir más rápido? No, seguir moviéndose con la misma *velocidad*. ¡Pero eso es lo que estamos tratando de definir! Porque si la pelota sigue moviéndose de la manera en que lo está haciendo, seguirá simplemente

moviéndose de la manera como lo está haciendo. Por lo tanto, necesitamos definir mejor la velocidad. ¿Qué debe seguir lo mismo? La dama puede también razonar de esta manera: "¡Si siguiera moviéndome de la manera como lo estoy haciendo durante una hora más, me iría contra esa pared al final de la calle!" No es tan fácil expresar lo que queremos decir.

Muchos físicos piensan que la medición es la única definición de cualquier cosa. Evidentemente, entonces, debiéramos usar el instrumento que mide la velocidad —el velocímetro— y decir: "Mire, señora, su velocímetro marca 100". Y entonces ella dice: "Mi velocímetro está roto y no marcaba en absoluto". ¿Significa esto que el auto está detenido? Creemos que hay algo que medir antes de construir el velocímetro. Sólo entonces podemos decir, por ejemplo: "El velocímetro no está funcionando bien", o "el velocímetro está roto". Ésa sería una frase sin sentido si la velocidad no tuviera un significado independiente del velocímetro. Así, pues, tenemos en nuestras mentes, evidentemente, una idea que es independiente del velocímetro, y el velocímetro está ideado sólo para medir esta idea. Por lo cual veamos si podemos obtener una mejor definición de la idea. Decimos: "Sí, por supuesto, antes de andar una hora usted chocaría esa muralla, pero si anduviera un segundo, recorrería 28 metros; señora, usted iba a 28 metros por segundo y si siguiera andando, el próximo segundo estaría a 28 metros y la muralla aquella está más lejos". Ella dice: "¡Sí, pero no hay ninguna ley que prohíba ir a 28 metros por segundo! Hay sólo una ley que prohíbe ir a 100 kilómetros por hora", "pero", replicamos "es la misma cosa". Si *es* la misma cosa, no sería necesario este circunloquio acerca de los 28 metros por segundo. En realidad, la pelota que está cayendo no puede seguir moviéndose de la misma manera ni siquiera un segundo, debido a que estaría cambiando la velocidad, y tendremos que definir la velocidad de alguna manera.

Ahora parece que estamos entrando en buen camino, es algo más o menos así: si la señora prosiguiera otro 1/1 000 de hora recorrería 1/1 000 de 100 kilómetros. En otras palabras, no tiene que proseguir la hora completa; la cuestión es que *por un momento* está yendo a esa velocidad. Ahora bien, lo que eso significa es que, si anduviera sólo un poquito más de tiempo, la distancia adicional que recorrería sería la misma que la de un auto que va a una velocidad *constante* de 100 kilómetros por hora. Tal vez la idea de 28 metros por segundo sea correcta: nosotros vemos hasta dónde llegó en el último segundo, se divide entre 28 metros y si sale 1 la velocidad era de 100 kilómetros por hora. En otras palabras, podemos encontrar la velocidad de esta manera: preguntamos, ¿qué distancia recorrimos en un tiempo muy corto? Dividimos esa distancia por el tiempo, y eso da la velocidad. Pero el tiempo debería hacerse tan corto como sea posible, cuanto más corto tanto mejor, porque algún cambio podría tener lugar durante este tiempo. Si tomamos el tiempo de un cuerpo que cae durante una hora, la idea es ridícula. Si lo tomamos durante un segundo, el resultado es bastante bueno para un auto, debido a que la velocidad no cambia mucho, pero no para un cuerpo que está cayendo; así, a fin de obtener la velocidad más y más exacta, deberíamos tomar intervalos de tiempo más y más cortos. Lo que deberíamos es tomar una millonésima de segundo, ver que tan lejos el auto se alejó y dividir esa distancia entre una millonésima de segundo. El resultado da la distancia entre segundo, lo cual es lo que entendemos por velocidad, por lo que podemos definirla de esa manera. Es una respuesta feliz a la señora, o más bien, es la definición que vamos a usar.

La definición anterior envuelve una nueva idea, una idea que los griegos no tenían en una forma general. Esa idea es pensar una *distancia infinitesimal* y el correspondiente *tiempo infinitesimal*, formar el cociente, y observar qué sucede a ese cociente cuando el tiempo que usamos llega a ser más y más corto. En otras palabras, tomar un límite de la distancia recorrida dividida entre el tiempo necesario, cuando el tiempo tomado es cada vez más pequeño, *hasta el infinito*. Esta idea fue inventada por Newton y Leibniz, de manera independiente, y es el comienzo de una nueva rama de las matemáticas, llamada *cálculo diferencial*. Éste fue inventado con el fin de describir el movimiento, y su primera aplicación fue al problema de definir qué significa ir "a 100 kilómetros por hora".

Tratemos de definir la velocidad un poco mejor. Supongamos que en un corto tiempo ϵ, el automóvil u otro cuerpo recorren una corta distancia x; entonces la velocidad, v, está definida por:

$$v = x/\epsilon,$$

una aproximación que mejora a medida que ϵ se hace más y más pequeño. Si deseamos una expresión matemática, podemos decir que la velocidad es igual al límite cuando ϵ se hace más y más pequeño en la expresión x/ϵ, o bien

$$v = \lim_{\epsilon \to 0} \frac{x}{\epsilon}. \tag{8.3}$$

No podemos hacer la misma cosa con la señora del automóvil, ya que la tabla es incompleta. Sólo sabemos dónde estaba a intervalos de un minuto; podemos obtener una idea aproximada de que iba a 1 700 m/min durante los primeros siete minutos, pero no sabemos exactamente si al cabo de siete minutos ella había estado aumentando la velocidad y que ésta era de 1 650 m/min al comienzo del sexto minuto y que ahora es 1 750 m/min, o algo así, debido a que no tenemos los detalles exactos del intermedio. Así, pues, sólo si la tabla se completara con un número infinito de datos podríamos realmente calcular la velocidad a partir de una tabla de esta forma. Por otro lado, cuando tenemos una fórmula matemática completa, como en el caso del cuerpo que cae (ecuación 8.1), entonces es posible calcular la velocidad, porque podremos calcular la posición en cualquier instante.

Tomemos como ejemplo el problema de determinar la velocidad, en el instante particular de cinco segundos, de una pelota que cae. Una manera de hacer esto es ver en la tabla 8-2 lo que hizo en los primeros cinco segundos; fue $125 - 80 = 45$ metros, así que iba a 45 m/s; sin embargo, eso es falso, debido a que la velocidad está cambiando; es *el promedio* de 45 m/s durante este intervalo, pero la pelota aumenta su velocidad y va realmente a más de 45 m/s. Queremos encontrar *exactamente a qué velocidad cae*. La técnica utilizada en este proceso es la siguiente: sabemos dónde se encuentra la pelota a los cinco segundos. A los 5.1 segundos, la distancia que ha caído es $5(5.1)^2 = 130.05$ metros (ver ecuación 8.1). A los cinco segundos ya ha caído 125 metros; en el último décimo de segundo cayó $130.05 - 125 = 5.05$ metros. Como 5.05 metros en 0.1 s es lo mismo que 50.5 m/s, ésa es más o menos la velocidad, pero no es exactamente la correcta. ¿Es la velocidad a los 5, a los 5.1, o la mitad del camino en 5.05 segundos, o bien, a qué instante *es* ésta velocidad? No importa —el problema es calcular la velocidad a los cinco segundos, y no tenemos exactamente eso; tenemos que hacer un trabajo mejor—. Así tomamos un milésimo de segundo más que cinco segundos, o sea 5.001 segundos, y calculamos la caída total según

$$s = 5(5.001)^2 = 5(25.010001) = 125.050005\,\text{m}.$$

En el último 0.001 s la pelota cayó 0.050005 m, y si dividimos ese número entre 0.001 s, obtenemos la velocidad de 50.005 m/s. Esto es más próximo, muy próximo, pero es todavía *inexacto*. Debería ser ahora evidente lo que debemos hacer para encontrar exactamente la velocidad. Para calcular esto matemáticamente planteamos el problema en forma un poco más abstracta, encontrar la velocidad en un tiempo especial t_0, que en el problema original eran 5 segundos. Ahora la distancia correspondiente a t_0, que llamamos s_0, es $5t_0^2$, o 125 metros en este caso. A fin de encontrar la velocidad, preguntamos: “En el instante t_0+ (un poquito), o $t_0 + \epsilon$, ¿dónde está el cuerpo?” La nueva posición es $5(t_0 + \epsilon)^2 = 5t_0^2 + 10t_0\epsilon + 5\epsilon^2$. Así está más lejos de lo que estaba antes, debido a que antes era sólo $5t_0^2$. Esta distancia la llamaremos s_0+ (un poquito más), o $s_0 + x$ (si x es un poco más). Si ahora restamos la distancia correspondiente a t_0 de la distancia correspondiente a $t_0 + \epsilon$, obtenemos x, la distancia adicional recorrida; $x = 10t_0\epsilon + 5\epsilon^2$. Nuestra primera aproximación para la velocidad es:

$$v = \frac{x}{\epsilon} = 10t_0 + 5\epsilon. \tag{8.4}$$

La verdadera velocidad es el valor de este cociente, x/ϵ, cuando ϵ se hace infinitamente pequeño. En otras palabras, después de formar el cociente, tomamos el

límite cuando ϵ se hace más y más pequeño, esto es, tiende a 0. La ecuación se reduce a:

$$v\,(\text{en el instante } t_0) = 10t_0.$$

En nuestro problema $t_0 = 5$ segundos, así que la solución es $v = 10 \times 5 = 50$ m/s. Unas pocas líneas arriba, donde tomamos a ϵ igual a 0.1 y 0.001 segundos sucesivamente, el valor que obtuvimos para v fue un poco mayor que éste, pero ahora vemos que la velocidad real es precisamente 50 m/s.

8-3 La velocidad como derivada

El procedimiento que acabamos de desarrollar se realiza tan a menudo en matemáticas, que por conveniencia se han asignado notaciones especiales a nuestras cantidades ϵ y x. La notación para ϵ, usado anteriormente, se convierte en Δt y x en Δs. Este Δt significa "un poco más de t", e implica que puede hacerse más pequeño. El prefijo Δ no es multiplicador, así como $\operatorname{sen}\theta$ no significa $s \cdot e \cdot n \cdot \theta$ —simplemente define un incremento de tiempo—, y nos recuerda su carácter especial. Δs tiene un significado análogo para la distancia s. Ya que Δ no es un factor, no puede ser simplificado en el cociente $\Delta s/\Delta t$ para dar s/t, así como la razón $\operatorname{sen}\theta/\operatorname{sen}2\theta$ no puede ser reducida a 1/2 por simplificación. En esta notación, la velocidad es igual al límite de $\Delta s/\Delta t$ cuando Δt se hace más y más pequeño, o bien:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}. \tag{8.5}$$

Esto es realmente lo mismo que nuestra expresión anterior (8.3) con ϵ y x; pero tiene la ventaja de mostrar que algo está cambiando, y nos indica qué es lo que cambia.

A propósito, con una buena aproximación tenemos otra ley, que dice que el cambio en la distancia de un punto en movimiento es la velocidad por el intervalo de tiempo, es decir, $\Delta s = v\Delta t$. Este enunciado es válido sólo si la velocidad no cambia durante este intervalo de tiempo, y esta condición es válida sólo en el límite cuando Δt tiende a 0. A los físicos les gusta escribirlo $ds = vdt$, ya que por dt ellos entienden como Δt cuando es muy pequeño; con esta interpretación la expresión es válida con gran aproximación. Si Δt es demasiado grande, la velocidad podría cambiar durante el intervalo, y la aproximación llegaría a ser menos exacta. Para un tiempo dt aproximándose a cero, se tiene precisamente que $ds = vdt$. Con esta notación podemos escribir (8.5) en la forma

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt}.$$

Tabla 8-3. Tabla abreviada de derivadas

s, u, v, w son funciones arbitrarias de t;
a, b, c, y n son constantes arbitrarias

Función	Derivada
$s = t^n$	$\dfrac{ds}{dt} = nt^{n-1}$
$s = cu$	$\dfrac{ds}{dt} = c\,\dfrac{du}{dt}$
$s = u + v + w + \cdots$	$\dfrac{ds}{dt} = \dfrac{du}{dt} + \dfrac{dv}{dt} + \dfrac{dw}{dt} + \cdots$
$s = c$	$\dfrac{ds}{dt} = 0$
$s = u^a v^b w^c \cdots$	$\dfrac{ds}{dt} = s\left(\dfrac{a}{u}\,\dfrac{du}{dt} + \dfrac{b}{v}\,\dfrac{dv}{dt} + \dfrac{c}{w}\,\dfrac{dw}{dt} + \cdots\right)$

La cantidad ds/dt, que encontramos arriba, se llama "derivada de s respecto a t" (este lenguaje ayuda a saber qué cosa cambia) y el complicado proceso de encontrarla se llama derivar, o derivación. Los ds y dt que aparecen separadamente se llaman *diferenciales*. Para familiarizarnos con las palabras, decimos que encontramos que la derivada de la función $5t^2$, o la derivada (respecto a t) de $5t^2$, es $10t$. Cuando nos acostumbramos a las palabras, las ideas se comprenden más fácilmente. Para practicar, encontremos la derivada de una función más complicada. Consideremos la fórmula $s = At^3 + Bt + C$, la cual puede describir el movimiento de un punto. Las letras A, B y C representan números constantes, como en la forma familiar general de una ecuación de segundo grado. Partiendo de la fórmula para el movimiento, queremos encontrar la velocidad en cualquier instante. Para encontrar la velocidad de la manera más elegante, cambiamos t por $t + \Delta t$ y notamos que s ha cambiado a $s+$ cierto Δs; entonces encontramos Δs en términos de Δt. Es decir:

$$\begin{aligned} s + \Delta s &= A(t + \Delta t)^3 + B(t + \Delta t) + C \\ &= At^3 + Bt + C + 3At^2\,\Delta t + B\,\Delta t + 3At(\Delta t)^2 + A(\Delta t)^3, \end{aligned}$$

pero ya que

$$s = At^3 + Bt + C,$$

encontramos que

$$\Delta s = 3At^2\,\Delta t + B\,\Delta t + 3At(\Delta t)^2 + A(\Delta t)^3.$$

Pero nosotros no queremos Δs —queremos Δs dividido entre Δt—. Dividimos la ecuación anterior entre Δt, obteniendo:

$$\frac{\Delta s}{\Delta t} = 3At^2 + B + 3At(\Delta t) + A(\Delta t)^2.$$

Cuando Δt tiende a 0, el límite de $\Delta s/\Delta t$ es ds/dt y es igual a

$$\frac{ds}{dt} = 3At^2 + B.$$

Éste es el proceso fundamental del cálculo diferencial, derivar funciones. El proceso es aún más simple de lo que parece. Observemos que cuando estos desarrollos contienen cualquier término con un cuadrado o un cubo o cualquier potencia más alta de Δt, tales términos pueden ser eliminados de inmediato, ya que tenderán a 0 cuando se tome el límite. Después de una pequeña práctica el proceso se hace más fácil, porque uno sabe lo que tiene que dejar de lado. Hay muchas reglas o fórmulas para derivar los diversos tipos de funciones. Pueden ser memorizadas, o pueden encontrarse en tablas. Una pequeña lista se encuentra en la tabla 8-3.

8-4 La distancia como una integral

Ahora tenemos que discutir el problema inverso. Supongamos que, en vez de una tabla de distancias, tenemos una tabla de velocidades en diferentes tiempos, partiendo desde cero. Para la pelota que cae, tales velocidades y tiempos se muestran en la tabla 8-4. Una tabla similar podría construirse para la velocidad del auto, anotando las lecturas del velocímetro cada minuto o cada medio minuto. Si sabemos a qué velocidad está yendo el coche en cualquier instante, ¿podemos determinar hasta dónde va? Este problema es justamente el inverso del resuelto anteriormente; se nos da la velocidad y se nos pide encontrar la distancia. ¿Cómo podemos encontrar la distancia si conocemos la velocidad? Si la velocidad del automóvil no es constante y la señora va a 100 kilómetros por hora por un momento, luego disminuye la velocidad, la aumenta, y así sucesivamente, ¿cómo podemos determinar hasta dónde ha ido? Esto es fácil. Usamos la misma idea y

Tabla 8-4

Velocidad de una pelota que cae

t (s)	v (m/s)
0	0
1	10
2	20
3	30
4	40

expresamos la distancia en términos de infinitesimales. Digamos: "En el primer segundo su velocidad fue tal y tal, y de la fórmula $\Delta s = v\Delta t$ podemos calcular hasta dónde fue el auto en el primer segundo a esa velocidad". Ahora, en el segundo siguiente su velocidad es aproximadamente la misma, pero ligeramente diferente, podemos calcular hasta dónde fue ella en el segundo siguiente tomando la nueva velocidad multiplicada por el tiempo. Procedemos en forma similar para cada segundo, hasta el final del viaje. Ahora tenemos un número de pequeñas distancias y la distancia total será la suma de todos estos pequeños pedazos.

Esto es, la distancia será la suma de las velocidades por los tiempos, es decir, $s = \sum v\ \Delta t$, donde la letra griega $\sum$ (sigma) es usada para indicar la suma. Para ser más preciso, es la suma de la velocidad en un cierto instante, digamos en el instante i, multiplicada por Δt.

$$s = \sum_i v(t_i)\,\Delta t. \tag{8.6}$$

La regla para los tiempos es que $t_{i+1} = t_i + \Delta t$. Sin embargo, la distancia que obtenemos por este método no será correcta, porque la velocidad cambia durante el intervalo de tiempo Δt. Si tomamos los tiempos suficientemente cortos, la suma es precisa, pues los tomamos más y más pequeños hasta obtener la exactitud deseada. El verdadero s es:

$$s = \lim_{\Delta t \to 0} \sum_i v(t_i)\,\Delta t. \tag{8.7}$$

Los matemáticos han inventado un símbolo para este límite, análogo al símbolo para la diferencial. La Δ se convierte en d para recordarnos que el tiempo es tan pequeño como pueda ser, la velocidad se llama entonces v en el instante t y la adición se escribe como una suma con una gran "s", $\int$ (del latín *summa*), que se ha distorsionado y ahora, por desgracia, se llama integral. Escribimos entonces

$$s = \int v(t)\,dt. \tag{8.8}$$

El proceso de sumar todos estos términos se llama integración y es el proceso inverso de la derivación. La derivada de esta integral es v, por lo que un operador (d) elimina el otro ($\int$). Uno puede obtener la fórmula para la integración tomando las fórmulas para las derivadas y aplicándolas en el sentido inverso, ya que están relacionadas inversamente entre sí. Así, uno puede construir su propia tabla de integrales derivando toda clase de funciones. Para cada fórmula con una derivada, obtenemos una fórmula de integral si la invertimos.

Cada función puede ser derivada analíticamente, es decir, el proceso se puede realizar algebraicamente, y conduce a una función definida. Pero no es posible de una manera simple escribir un valor analítico para cualquier integral que se quiera. Pueden calcularlo, por ejemplo, haciendo la suma anterior, y luego haciéndola otra vez con un intervalo más fino Δt y otra vez con un intervalo más fino, hasta que estén cerca de lo correcto. En general, dada alguna función particular, no es posible encontrar, analíticamente, cuánto vale la integral. Uno puede siempre tratar de encontrar una función que cuando se deriva da alguna función deseada, pero podría no encontrarla y podría no existir, en el sentido de ser expresable en términos de funciones a las cuales ya se les ha dado nombre.

8-5 Aceleración

El próximo paso en el desarrollo de las ecuaciones de movimiento es introducir otra idea que va más allá del concepto de velocidad, al de *cambio* de velocidad, y ahora nos preguntamos: "¿Cómo *cambia* la velocidad?" En los capítulos anteriores hemos discutido casos en los cuales las fuerzas producen cambios de velocidad. Puede que hayan oído con gran interés acerca de algún automóvil que puede alcanzar desde el reposo 100 kilómetros por hora en 10 segundos en el plano. A partir de este comportamiento podemos ver con qué rapidez cambia

la velocidad, pero sólo en promedio. Lo que ahora discutiremos es el próximo nivel de complejidad, lo cual es con qué rapidez está cambiando la velocidad. En otras palabras, ¿cuántos metros por segundo cambia la velocidad en un segundo, esto es, cuántos metros por segundo, por segundo? Anteriormente dedujimos la fórmula para la velocidad de un cuerpo que cae como $v = 10t$, la cual está indicada en la tabla 8-4, y ahora queremos encontrar cuánto cambia la velocidad cada segundo, esta cantidad se llama aceleración.

La aceleración se define como la rapidez de cambio de la velocidad en el tiempo. De la discusión anterior ya sabemos lo suficiente para escribir la aceleración como la derivada dv/dt, de la misma manera que la velocidad es la derivada de la distancia. Si ahora derivamos la fórmula $v = 10t$ obtenemos para un cuerpo que cae,

$$a = \frac{dv}{dt} = 10. \tag{8.9}$$

[Para derivar el término $10t$ podemos utilizar el resultado obtenido en un problema anterior, donde encontraremos que la derivada de Bt es simplemente B (una constante). Así, si hacemos $B = 10$, tenemos inmediatamente que la derivada de $10t$ es 10.] Esto significa que la velocidad de un cuerpo que cae está cambiando siempre a razón de 10 metros por segundo. También vemos en la tabla 8-4 que la velocidad aumenta en 10 metros cada segundo. Éste es un caso muy simple, porque las aceleraciones usualmente no son constantes. La razón de que la aceleración es constante aquí es que la fuerza sobre el cuerpo que cae es constante, y la ley de Newton dice que la aceleración es proporcional a la fuerza.

Como un ejemplo más, encontremos la aceleración en el problema que ya hemos resuelto para la velocidad. Partiendo con:

$$s = At^3 + Bt + C$$

Obtuvimos, para $v = ds/dt$,

$$v = 3At^2 + B.$$

Como la aceleración es la derivada de la velocidad respecto al tiempo, debemos derivar la última expresión. Recordemos la regla de la derivada de los dos términos del segundo miembro, ésta es igual a la suma de las derivadas de los términos individualmente. Para derivar el primero de estos términos, en vez de ir a través del proceso fundamental, notemos otra vez que ya hemos derivado un término cuadrático cuando derivamos $5t^2$ y el efecto fue duplicar el coeficiente numérico y cambiar el t^2 por t; supongamos que la misma cosa sucederá esta vez, y ustedes pueden comprobar el resultado. La derivada de $3At^2$ será $6At$. Enseguida derivamos B, un término constante; pero, por la regla establecida previamente, la derivada de B es cero; por lo tanto, este término no contribuye a la aceleración. El resultado final es, por lo tanto, $a = dv/dt = 6At$.

A título de referencia, establezcamos dos fórmulas muy útiles, que se pueden obtener por integración. Si un cuerpo parte del reposo y se mueve con una aceleración constante g, su velocidad v en cualquier instante t está dada por:

$$v = gt.$$

La distancia que recorre en el mismo tiempo es:

$$s = \tfrac{1}{2}gt^2.$$

Hay varias notaciones matemáticas para escribir derivadas. Ya que la velocidad es ds/dt y la aceleración es la derivada de la velocidad respecto al tiempo, podemos también escribir:

$$a = \frac{d}{dt}\left(\frac{ds}{dt}\right) = \frac{d^2s}{dt^2}, \tag{8.10}$$

que son maneras comunes de escribir una segunda derivada.

Tenemos otra ley en que la velocidad es igual a la integral de la aceleración. Esto es justamente lo inverso de $a = dv/dt$; ya hemos visto que la distancia es la

integral de la velocidad, por lo que se puede encontrar la distancia por una doble integración de la aceleración.

En la discusión anterior el movimiento fue en una sola dimensión, y por razones de espacio sólo permiten una breve discusión del movimiento en tres dimensiones. Consideremos una partícula P, que se mueve en tres dimensiones de cualquier manera. Al comienzo de este capítulo, abrimos nuestra discusión del caso unidimensional de un automóvil en movimiento, observando la distancia recorrida por el automóvil desde su punto de partida en varios instantes. Luego se discutió la velocidad en términos de los cambios de esta distancia con el tiempo y la aceleración en términos de los cambios de la velocidad. Podemos tratar el movimiento tridimensional en forma análoga. Será más simple ilustrar el movimiento en un diagrama bidimensional y luego extender las ideas a tres dimensiones. Establezcamos un par de ejes en ángulo recto y determinemos la posición de la partícula en cualquier instante midiendo a qué distancia está de cada uno de los dos ejes. Así cada posición está dada en términos de una distancia x y una distancia y, y el movimiento puede ser descrito construyendo una tabla en la cual estas distancias están dadas en función del tiempo. (La extensión de este proceso a tres dimensiones sólo requiere otro eje, perpendicular a los dos primeros, y medir una tercerea distancia, la distancia z. Las distancias se miden ahora desde *planos* coordenados en vez de líneas.) Habiendo construido una tabla con las distancias x y y, ¿cómo podemos determinar la velocidad? Primero encontramos las componentes de la velocidad en cada dirección. La parte horizontal de la velocidad, o componente x, es la derivada de la distancia x respecto al tiempo, o bien:

$$v_x = dx/dt. \tag{8.11}$$

En forma similar, la parte vertical de la velocidad, o componente y, es

$$v_y = dy/dt. \tag{8.12}$$

En la tercera dimensión,

$$v_z = dz/dt. \tag{8.13}$$

Ahora bien, dadas las componentes de la velocidad, ¿cómo podemos encontrar la velocidad a lo largo de la trayectoria real del movimiento? En el caso bidimensional, consideremos dos posiciones sucesivas de la partícula, separadas por una distancia corta Δs y un pequeño intervalo de tiempo $t_2 - t_1 = \delta t$. En el tiempo Δt la partícula se mueve horizontalmente una distancia $\Delta x \approx v_x \Delta t$ y verticalmente una distancia $\Delta y \approx v_y \Delta t$. (El símbolo "$\approx$" se lee "es aproximadamente".) La distancia real que se mueve es aproximadamente:

$$\Delta s \approx \sqrt{(\Delta x)^2 + (\Delta y)^2}, \tag{8.14}$$

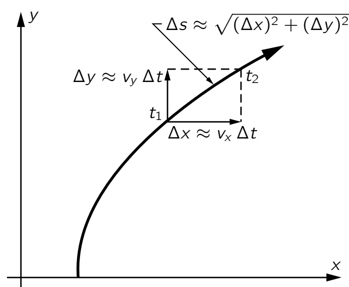


Figura 8-3. Descripción, en dos dimensiones, del movimiento de un cuerpo y el cálculo de su velocidad.

como se muestra en la figura 8-3. La velocidad aproximada durante este intervalo puede obtenerse dividiendo por Δt y haciendo que Δt tienda a 0, como al comienzo del capítulo. Entonces, obtenemos la velocidad como:

$$v = \frac{ds}{dt} = \sqrt{(dx/dt)^2 + (dy/dt)^2} = \sqrt{v_x^2 + v_y^2}. \tag{8.15}$$

Para tres dimensiones el resultado es:

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{8.16}$$

De la misma manera como definimos las velocidades, podemos definir las aceleraciones, tenemos una componente x de la aceleración a_x, que es la derivada de v_x, la componente x de la velocidad ($a_x = d^2x/dt^2$ es la segunda derivada de x respecto a t), y así sucesivamente.

Consideremos un bonito ejemplo de un movimiento compuesto en un plano. Tomaremos un movimiento en el cual una pelota se mueve horizontalmente con una

velocidad constante u y al mismo tiempo cae verticalmente con una aceleración constante $-g$. ¿Cuál es el movimiento? Podemos decir que $dx/dt = v_x = u$. Ya que la velocidad v_x es constante,

$$x = ut, \tag{8.17}$$

y ya que la aceleración hacia abajo $-g$ es constante, la distancia y que el objeto cae puede describirse en la forma:

$$y = -\tfrac{1}{2}gt^2. \tag{8.18}$$

¿Cuál es la curva de su trayectoria, es decir, cuál es la relación entre y y x? Podemos eliminar t de la ecuación (8.18), ya que $t = x/u$. Haciendo esta sustitución, encontramos que:

$$y = -\frac{g}{2u^2}\,x^2. \tag{8.19}$$

Esta relación entre y y x puede ser considerada como la ecuación de la trayectoria de la pelota que se mueve. Dibujando esta ecuación, obtenemos una curva que se llama parábola. Todo cuerpo que cae libremente disparado en cualquier dirección describirá una parábola, como se muestra en la figura 8-4.

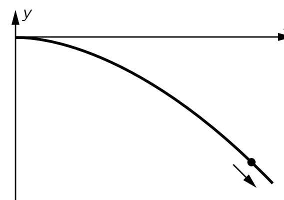


Figura 8-4. La parábola descrita por un cuerpo que cae con una velocidad horizontal inicial.

9

Leyes de Newton de la dinámica

9-1 Momentum y fuerza

El descubrimiento de las leyes de la dinámica, o de las leyes del movimiento, fue un momento culminante en la historia de la ciencia. Antes de la época de Newton, los movimientos de las cosas como los planetas eran un misterio, pero después de Newton hubo una comprensión completa. Aun las leves variantes de las Leyes de Kepler, debido a las perturbaciones de los planetas, fueron computables. El movimiento de los péndulos, osciladores con resortes y pesos en ellos, etc., pudieron ser analizados completamente después de que las leyes de Newton fueron enunciadas. Antes de este capítulo no pudimos calcular cómo se movería una masa fija a un resorte, mucho menos podríamos calcular las perturbaciones del planeta Urano debido a Júpiter y Saturno. ¡Después de este capítulo *podremos* calcular no sólo el movimiento de las masas que oscilan, sino también las perturbaciones producidas por Júpiter y Saturno sobre el planeta Urano!

Galileo hizo un gran avance en la comprensión del movimiento cuando descubrió el *principio de inercia*, si un objeto se abandona y si no es perturbado, continúa moviéndose con una velocidad constante en una línea recta, si estaba originalmente moviéndose; o continúa en reposo, si estaba en reposo. Por supuesto, éste nunca parece ser el caso en la naturaleza, porque si hacemos deslizar un bloque a lo largo de una mesa se detiene, pero eso es debido a que *no* ha sido abandonado —está rozando contra la mesa—. Requirió cierta imaginación encontrar la regla correcta y esa imaginación la tuvo Galileo.

Por supuesto, la próxima cosa que se necesita es una regla para encontrar cómo un objeto *cambia* su velocidad si algo lo *está* afectando. Ésta es la contribución de Newton. Escribió tres leyes: la primera ley fue una mera reafirmación del principio de inercia galileano recién descrito. La segunda ley dio una manera específica para determinar cómo la velocidad cambia bajo diferentes influencias llamadas *fuerzas.* La tercera ley describe las fuerzas con algún detalle y discutiremos eso en otra oportunidad. Aquí discutiremos sólo la segunda ley, la cual sostiene que las fuerzas cambian el movimiento de un objeto, de este modo: *la variación temporal de una cantidad llamada momentum es proporcional a la fuerza.* Enunciaremos pronto esto en forma matemática, pero expliquemos primero la idea.

El *momentum* no es lo mismo que la *velocidad.* Muchas palabras se usan en física, y todas tienen significado preciso en ella, aunque no puedan tener tal significado preciso en el lenguaje cotidiano. El momentum es un ejemplo y debemos definirlo en forma precisa. Si ejercemos un cierto empuje con nuestros brazos sobre un objeto liviano, éste se mueve fácilmente; si empujamos igualmente fuerte otro objeto mucho más pesado en el sentido usual, se mueve mucho menos rápido. Realmente, debemos cambiar las palabras "liviano" y "pesado" a *menos masivo* y *más masivo*, porque hay una diferencia que debe entenderse entre el *peso* de un objeto y su *inercia.* (Lo difícil que es poner una cosa en movimiento y cuánto pesa es algo diferente.) El peso y la inercia son *proporcionales* y sobre la superficie de la Tierra se consideran a menudo numéricamente iguales, lo cual causa una cierta confusión al estudiante. Sobre Marte, los pesos serían diferentes, pero la magnitud de la fuerza necesaria para vencer la inercia sería la misma.

Usamos el término *masa* como medida cuantitativa de la inercia, y podemos medir masa, por ejemplo, haciendo girar un objeto en círculo a determinada velocidad y midiendo cuánta fuerza necesitamos para mantenerlo en círculo. De esta manera, encontramos cierta cantidad de masa para cada objeto. Ahora, el

momentum de un objeto es un producto de dos partes: su *masa* y su *velocidad.* Así, pues, la segunda ley de Newton puede escribirse matemáticamente de esta manera:

$$F = \frac{d}{dt}(mv). \tag{9.1}$$

Ahora hay varios puntos a considerar. Al escribir cualquier ley como ésta, usamos muchas ideas intuitivas, implicaciones y suposiciones, las cuales primero se combinan aproximadamente en nuestra "ley". Posteriormente, tenemos que volver atrás y estudiar con mayor detalle lo que cada uno de los términos significa exactamente, pero si tratamos de hacerlo demasiado pronto nos confundiremos. Así, pues, al comienzo daremos varias cosas por supuestas. Primero, la masa de un objeto es *constante*; no lo es realmente, pero partiremos con la aproximación newtoniana de que la masa es constante, la misma todo el tiempo, y que, además, cuando ponemos dos objetos juntos, sus masas se *suman.* Estas ideas fueron por cierto insinuadas por Newton cuando escribió su ecuación, pues de otro modo ésta no tiene sentido. Por ejemplo, supongamos que la masa variará inversamente con la velocidad; entonces, el momentum *no cambiaría nunca* en circunstancia alguna, así la ley no significa nada a menos que sepan cómo cambia la masa respecto a la velocidad. Para comenzar a decirnos *que no cambia.*

Luego hay algunas implicaciones concernientes a la fuerza. Como una aproximación, imaginamos la fuerza como una especie de empuje o tirón que hacemos con nuestros músculos, pero podemos definirla con mayor exactitud pues ahora tenemos esta ley del movimiento. La cosa más importante de reconocer es que esta relación comprende no sólo el cambio en la *magnitud* del *momentum* o de la velocidad, sino que también en su *dirección.* Si la masa es constante, la ecuación (9.1) puede también escribirse en la forma:

$$F = m\,\frac{dv}{dt} = ma. \tag{9.2}$$

La aceleración a es la variación de la velocidad, y la segunda ley de Newton no dice sólo que el efecto de una fuerza dada varía inversamente con la masa; dice también que la *dirección* del cambio de la velocidad y la *dirección* de la fuerza son iguales. Así debemos comprender que un cambio de velocidad, o una aceleración, tiene un significado más amplio que en el lenguaje común, la velocidad de un objeto en movimiento puede cambiar porque se mueve más rápido, o más lento por su disminución (en este caso, decimos que acelera con una aceleración negativa), o cambiando su dirección de movimiento. Una aceleración en ángulo recto con la velocidad fue discutida en el capítulo 9. Vimos allí que un objeto que se mueve en un círculo de radio R con una cierta velocidad v a lo largo del círculo se desvía de una trayectoria rectilínea en una distancia igual a $\frac{1}{2}(v^2/R)t^2$, si t es muy pequeño. Así, la fórmula para la aceleración en ángulo recto al movimiento es:

$$a = v^2/R, \tag{9.3}$$

y una fuerza en ángulo recto a la velocidad obligaría a un objeto a moverse en una trayectoria curva cuyo radio de curvatura se puede encontrar dividiendo la fuerza entre la masa para obtener la aceleración, y usando luego la ecuación (9.3).

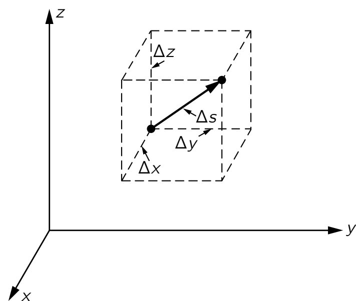


Figura 9-1. Un pequeño desplazamiento de un objeto.

9-2 La velocidad tiene dirección

Con el propósito de hacer más preciso nuestro lenguaje, haremos una definición más en nuestro uso de la palabra *velocidad.* Ordinariamente pensamos que rapidez y velocidad son lo mismo, y en lenguaje ordinario ellas *son* lo mismo. Podemos formular esto más precisamente describiendo cómo cambian las coordenadas x, y y z de un objeto con el tiempo. Supongamos, por ejemplo, que en un cierto instante un objeto se está moviendo como se muestra en la figura 9-1. En un

pequeño intervalo de tiempo dado Δt se moverá en una cierta distancia Δx en la dirección x, Δy en la dirección y y Δz en la dirección z. El efecto total de estos tres cambios en las coordenadas es un desplazamiento Δs a lo largo de la diagonal de un paralelepípedo cuyos lados son Δx, Δy y Δz. En términos de la velocidad, el desplazamiento Δx es la componente x de la velocidad por Δt y similarmente para Δy y Δz:

$$\Delta x = v_x\,\Delta t, \qquad \Delta y = v_y\,\Delta t, \qquad \Delta z = v_z\,\Delta t. \tag{9.4}$$

9-3 Componentes de la velocidad, de la aceleración y de la fuerza

En la ecuación (9.4) *hemos descompuesto la velocidad en componentes* diciendo lo rápido que el objeto se está moviendo en la dirección x, la dirección y y la dirección z. La velocidad está completamente especificada, en módulo y dirección, si damos los valores numéricos de sus tres componentes cartesianas:

$$v_x = dx/dt, \qquad v_y = dy/dt, \qquad v_z = dz/dt. \tag{9.5}$$

Por otro lado, el módulo de la velocidad del objeto es

$$ds/dt = |v| = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{9.6}$$

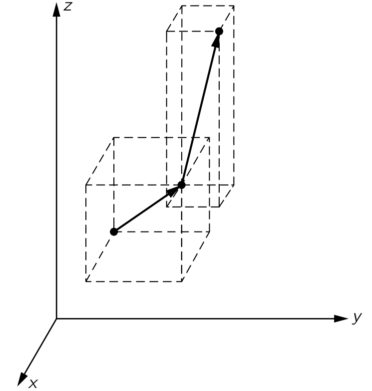


Figura 9-2. Un cambio en la velocidad en el cual su magnitud y dirección cambian.

A continuación, supongamos que, debido a la acción de una fuerza, la velocidad cambia a alguna otra dirección y a una magnitud diferente, como se muestra en la figura 9-2. Podemos analizar esta situación aparentemente compleja en forma más bien simple si calculamos la variación de las componentes x, y y z de la velocidad. La variación de la componente de la velocidad en la dirección x en un tiempo Δt es $\Delta v_x = a_x \Delta t$, donde a_x es lo que llamamos la componente x de la aceleración. Similarmente, vemos que $\Delta v_y = a_y \Delta t$ y $\Delta v_z = a_z \Delta t$. En estos términos vemos que la segunda ley de Newton, al decir que la fuerza está en la misma dirección que la aceleración, es realmente tres leyes, en el sentido de que la componente de la fuerza en la dirección x, y o z es igual a la masa por la variación de la correspondiente componente de la velocidad:

$$\begin{aligned} F_x &= m(dv_x/dt) = m(d^2x/dt^2) = ma_x, \\ F_y &= m(dv_y/dt) = m(d^2y/dt^2) = ma_y, \\ F_z &= m(dv_z/dt) = m(d^2z/dt^2) = ma_z. \end{aligned} \tag{9.7}$$

Así como la velocidad y la aceleración han sido descompuestas en componentes proyectando un segmento de recta que representa la cantidad y dirección sobre tres ejes coordenados, así, de la misma manera, una fuerza en una dirección dada se representa por ciertas componentes en las direcciones x, y y z:

$$\begin{aligned} F_x &= F\cos(x, F), \\ F_y &= F\cos(y, F), \\ F_z &= F\cos(z, F), \end{aligned} \tag{9.8}$$

donde F es el módulo (magnitud) de la fuerza y (x, F) representa el ángulo entre el eje x y la dirección de F, etcétera.

La segunda ley de Newton está dada en su forma completa en la ecuación (9.7). Si conocemos las fuerzas que actúan sobre un objeto y las resolvemos en componentes x, y y z, entonces podemos encontrar el movimiento del objeto a partir de estas ecuaciones. Consideremos un ejemplo simple. Supongamos que no hay fuerza en las direcciones y y z, la única fuerza está en la dirección x, digamos verticalmente. La ecuación (9.7) nos dice que habrá cambios en la velocidad en la dirección vertical, pero no habrá cambio en las direcciones horizontales. Esto fue demostrado con un aparato especial en el capítulo 7 (figura 7-3). Un cuerpo que cae se mueve horizontalmente sin cambiar su movimiento horizontal, mientras que se mueve verticalmente de la misma manera como se movería si el movimiento horizontal fuera cero. En otras palabras, los movimientos en las direcciones x, y y z son independientes si las *fuerzas* no están relacionadas.

9-4 ¿Cuál es la fuerza?

A fin de usar las leyes de Newton, debemos contar con algunas fórmulas para la fuerza; estas leyes dicen: *pongan atención a las fuerzas.* Si un objeto está acelerando, algún agente está trabajando, encuéntrelo. Nuestro programa para el futuro de la dinámica debe ser *encontrar las leyes de la fuerza.* Newton mismo llegó a dar algunos ejemplos. En el caso de la gravedad dio una forma específica para la fuerza. En el caso de otras fuerzas dio parte de la información en su tercera ley, que se estudiará en el próximo capítulo y que tiene que ver con la igualdad de la acción y la reacción.

Extendiendo nuestro ejemplo anterior, ¿cuáles son las fuerzas sobre los objetos cerca de la superficie de la Tierra? Cerca de la superficie de la Tierra, la fuerza en la dirección vertical debido a la gravedad es proporcional a la masa del objeto y es aproximadamente independiente de la altura para alturas pequeñas comparadas con el radio de la Tierra $R: F = GmM/R^2 = mg$, donde $g = GM/R^2$ se llama *aceleración gravitacional.* Así pues, la ley de la gravedad nos dice que el peso es proporcional a la masa, la fuerza está en la dirección vertical y es la masa multiplicada por g. Otra vez encontramos que el movimiento en la dirección horizontal es una velocidad constante. El movimiento interesante está en la dirección vertical, y la segunda ley de Newton nos dice:

$$mg = m(d^2x/dt^2). \tag{9.9}$$

Simplificando las m, encontramos que la aceleración en la dirección x es constante e igual a g. Ésta es, por supuesto, la bien conocida ley de la caída libre bajo la acción de la gravedad, la cual conduce a las ecuaciones,

$$\begin{aligned} v_x &= v_0 + gt, \\ x &= x_0 + v_0t + \tfrac{1}{2}gt^2. \end{aligned} \tag{9.10}$$

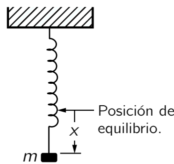


Figura 9-3. Una masa fija a un resorte.

Como otro ejemplo, supongamos que hemos podido construir un artefacto (figura 9-3) que aplica una fuerza proporcional a la distancia y dirigida en forma opuesta —un resorte—. Si olvidamos la gravedad, la cual por supuesto está balanceada por el estiramiento inicial del resorte y hablamos sólo de las fuerzas *en exceso*, vemos que si tiramos la masa hacia abajo, el resorte tira hacia arriba, mientras que si la empujamos hacia arriba el resorte empuja hacia abajo. Esta máquina ha sido diseñada cuidadosamente de modo que la fuerza es tanto mayor mientras más tiramos de ella, en proporción exacta al desplazamiento desde la condición de equilibrio, y la fuerza hacia arriba es igualmente proporcional a cuanto tiramos hacia abajo. Si observamos la dinámica de esta máquina, vemos un movimiento bastante hermoso —arriba, abajo, arriba, abajo,. . . —. La pregunta es, ¿describirán correctamente las ecuaciones de Newton este movimiento? Veamos si podemos calcular exactamente cómo se mueve con esta oscilación periódicas, aplicando la ley de Newton (9.7). En el caso presente, la ecuación es:

$$-kx = m(dv_x/dt). \tag{9.11}$$

Aquí tenemos una situación donde la velocidad en la dirección x cambia proporcionalmente a x. Nada se ganará teniendo numerosas constantes, así que imaginaremos que la escala de tiempo ha cambiado o que hay un accidente en las unidades, de modo que tengamos $k/m = 1$. Así trataremos de resolver la ecuación

$$dv_x/dt = -x. \tag{9.12}$$

Para proseguir debemos saber qué es v_x; pero, por supuesto, sabemos que la velocidad es la variación de la posición con el tiempo.

9-5 Significado de las ecuaciones de la dinámica

Tratemos ahora de analizar lo que significa la ecuación (9.12). Supongamos que en un tiempo dado t el objeto tiene una cierta velocidad v_x y posición x. ¿Cuál es la velocidad y cuál es la posición en un tiempo $t+\epsilon$ inmediatamente posterior? Si podemos responder a esta pregunta, nuestro problema está resuelto, porque entonces podemos comenzar con las condiciones dadas y calcular cómo cambian para el primer instante, el instante siguiente y el instante siguiente, y así sucesivamente y de esta manera gradualmente desarrollamos el movimiento. Para fijar ideas, supongamos que en el tiempo $t=0$ tenemos que $x=1$ y $v_x=0$. ¿Por qué se mueve el objeto? Porque hay una *fuerza* sobre él cuando está en cualquier posición, excepto en $x=0$. Si $x>0$, la fuerza está hacia arriba. Por lo tanto, la velocidad que es cero comienza a cambiar, debido a la ley del movimiento. Una vez que comienza a establecerse cierta velocidad, el objeto empieza a moverse hacia arriba, y así sucesivamente. Ahora bien, en cualquier instante $t+\epsilon$ es muy pequeño, podemos expresar la posición en el instante $t+\epsilon$ en términos de la posición en el instante t y la velocidad en el instante t con una muy buena aproximación como:

$$x(t+\epsilon)=x(t)+\epsilon v_x(t). \tag{9.13}$$

Cuanto más pequeño es el ϵ, tanto con mayor exactitud es la expresión, pero es de una exactitud útil aún si ϵ no es suficientemente pequeño. Ahora bien ¿qué pasa con la velocidad? A fin de obtener la velocidad posterior, la velocidad en el instante $t+\epsilon$, necesitamos conocer cómo cambia la velocidad, la *aceleración*. ¿Y cómo vamos a encontrar la aceleración? Aquí es donde la ley de la dinámica interviene. La ley de la dinámica nos dice lo que es la aceleración. Nos dice que la aceleración es $-x$,

$$v_x(t+\epsilon)=v_x(t)+\epsilon a_x(t) \tag{9.14}$$

$$=v_x(t)-\epsilon x(t). \tag{9.15}$$

La ecuación (9.14) es meramente cinemática, nos dice que una velocidad cambia debido a la presencia de la aceleración. Pero la ecuación (9.15) es *dinámica*, porque nos relaciona la aceleración con la fuerza, nos dice que en este instante particular, para este problema particular, pueden remplazar la aceleración por $-x(t)$. Por lo tanto, si conocemos la x y la v en el instante dado, conocemos la aceleración, la cual nos da la nueva velocidad, y con ésta la nueva posición —y así es como trabaja el método—. La velocidad cambia un poquito debido a la fuerza, y la posición cambia un poquito debido a la velocidad.

9-6 Solución numérica de las ecuaciones

Resolvamos ahora realmente el problema. Supongamos que tomamos $\epsilon=0.100$ segundos. Después que hacemos todo el trabajo y si encontramos que éste no es suficientemente pequeño, tendríamos que volver atrás y hacerlo de nuevo con $\epsilon=0.010$ segundos. Comenzando con nuestro valor inicial $x(O)=1.00$, ¿cuánto es $x(0.1)$? Es la posición antigua $x(0)$ más la velocidad (que es cero) multiplicada por 0.10 segundos. Así, $x(0.1)$ es aún 1.00 porque aún no ha comenzado a moverse. Pero la nueva velocidad a los 0.10 segundos será la velocidad antigua $v(0)=0$ más ϵ por la aceleración. La aceleración es $-x(0)=-1.00$. Entonces

$$v(0.1)=0.00-0.10\times 1.00=-0.10.$$

Ahora a 0.20 segundos,

$$\begin{aligned}x(0.2)&=x(0.1)+\epsilon v(0.1)\\&=1.00-0.10\times 0.10=0.99\end{aligned}$$

y

$$\begin{aligned}v(0.2)&=v(0.1)+\epsilon a(0.1)\\&=-0.10-0.10\times 1.00=-0.20.\end{aligned}$$

Tabla 9-1

Solución de $dv_x/dt = -x$
Intervalo: $\epsilon = 0.10$ s

t	x	v_x	a_x
0.0	1.000	0.000	−1.000
		−0.050	
0.1	0.995		−0.995
		−0.150	
0.2	0.980		−0.980
		−0.248	
0.3	0.955		−0.955
		−0.343	
0.4	0.921		−0.921
		−0.435	
0.5	0.877		−0.877
		−0.523	
0.6	0.825		−0.825
		−0.605	
0.7	0.764		−0.764
		−0.682	
0.8	0.696		−0.696
		−0.751	
0.9	0.621		−0.621
		−0.814	
1.0	0.540		−0.540
		−0.868	
1.1	0.453		−0.453
		−0.913	
1.2	0.362		−0.362
		−0.949	
1.3	0.267		−0.267
		−0.976	
1.4	0.169		−0.169
		−0.993	
1.5	0.070		−0.070
		−1.000	
1.6	−0.030		+0.030

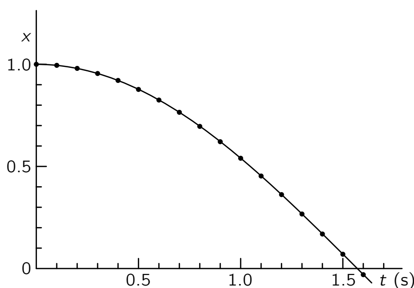


Figura 9-4. Gráfico del movimiento de una masa fija a un resorte.

y así, una y otra vez más, podemos calcular el resto del movimiento y esto es justamente lo que haremos. Sin embargo, para fines prácticos, hay algunos pequeños trucos con los cuales podemos aumentar la exactitud. Si continuáramos este cálculo como hemos comenzado, encontraríamos que el movimiento es bastante inexacto, ya que $\epsilon = 0.100$ segundos es bastante grande y tendríamos que ir a un intervalo muy pequeño, $\epsilon = 0.01$ digamos. Entonces para pasar por un intervalo de tiempo total razonable tomaría muchos ciclos de computación. Por eso organizaremos el trabajo de una manera que aumentará la precisión de nuestros cálculos, usando el mismo tosco intervalo de $\epsilon = 0.10$ segundos. Esto puede realizarse si hacemos un mejoramiento sutil en la técnica del análisis.

Noten que la nueva posición es la posición antigua más el intervalo de tiempo ϵ por la velocidad. Pero por la velocidad *¿cuándo*? La velocidad al comienzo del intervalo de tiempo es una velocidad y la velocidad al final del intervalo de tiempo es otra velocidad. Nuestra mejora es usar la velocidad *media entre*. Si conocemos la velocidad ahora, pero la velocidad está cambiando, no obtendremos la respuesta correcta al ir a la misma como ahora. Deberíamos usar alguna entre la rapidez "ahora" y la "después" al final del intervalo. Las mismas consideraciones se aplican también a la velocidad; para calcular los cambios de velocidad deberíamos usar la aceleración a mitad del intervalo entre los dos tiempos a los cuales la velocidad debe ser encontrada. Así, las ecuaciones que realmente usaremos serán algo como esto: la posición posterior es igual a la posición anterior más ϵ por la velocidad *en el instante a mitad del intervalo*. En forma similar, la velocidad en este punto medio es la velocidad en el tiempo ϵ anterior (que está en la mitad del intervalo anterior), más ϵ por la aceleración en el instante t. Esto es, usamos las ecuaciones:

$$\begin{aligned} x(t+\epsilon) &= x(t) + \epsilon v(t+\epsilon/2), \\ v(t+\epsilon/2) &= v(t-\epsilon/2) + \epsilon a(t), \\ a(t) &= -x(t). \end{aligned} \tag{9.16}$$

Aquí queda sólo un pequeño problema, ¿cuánto vale $v(\epsilon/2)$? Al comienzo tenemos $v(0)$, no $v(-\epsilon/2)$. Para iniciar nuestro cálculo, usaremos una ecuación especial, es decir, $v(\epsilon/2) = v(0) + (\epsilon/2)a(0)$.

Ahora estamos listos para llevar a cabo nuestro cálculo. Por conveniencia podemos disponer el trabajo en forma de tabla, con columnas para el tiempo, la posición, la velocidad y la aceleración, y las líneas intermedias para la velocidad, como se muestra en la tabla 9-1. Tal tabla es, por supuesto, una manera conveniente de representar los valores numéricos obtenidos a partir del conjunto de ecuaciones (9.16), y de hecho no es necesario escribir las ecuaciones mismas. Llenamos simplemente los diversos espacios de la tabla uno por uno. Esta tabla nos da ahora una idea muy buena del movimiento: comienza desde el reposo, primero toma una pequeña velocidad hacia arriba (negativa) y pierde algo de su distancia. La aceleración es entonces un poquito menos pequeña, pero aún está ganando velocidad. Pero al seguir va ganando velocidad más y más lentamente; hasta que pasa $x = 0$ alrededor de $t = 1.50$ segundos podemos predecir confiadamente que se seguirá moviendo, pero ahora estará en el otro lado, la posición x llegará a ser negativa, y por lo tanto la aceleración será positiva. Así, pues, la velocidad decrece. Es interesante comparar estos números con la función $x = \cos t$, que se da en la figura 9-4. ¡La concordancia está dentro de las tres cifras significativas de exactitud de nuestro cálculo! Veremos posteriormente que $x = \cos t$ es la solución matemática exacta de nuestra ecuación de movimiento, pero es una ilustración impresionante del poder del análisis numérico que un cálculo tan fácil dé resultados tan precisos.

9-7 Movimientos planetarios

El análisis anterior es muy bonito para el movimiento de un resorte que oscila, pero ¿podemos analizar el movimiento de un planeta alrededor del Sol? Veamos si podemos llegar a la aproximación de una elipse para la órbita. Supondremos que el Sol es infinitamente pesado, en el sentido de que no incluiremos su movimiento.

Supongamos que un planeta parte de un cierto lugar y que se está moviendo con una cierta velocidad, gira alrededor del Sol siguiendo alguna curva, y trataremos de analizar, mediante las leyes de Newton del movimiento y su ley de la gravitación, qué curva es. ¿Cómo? En un instante dado está en algún lugar en el espacio. Si la distancia radial desde el Sol a este lugar se llama r, sabemos que hay una fuerza dirigida hacia adentro que, de acuerdo a la ley de la gravedad, es igual a una constante por el producto de las masas del Sol y del planeta, dividido entre el cuadrado de la distancia. Para analizar más aún esto debemos encontrar qué aceleración producirá esta fuerza. Nosotros necesitaremos las *componentes* de la aceleración en dos direcciones, que llamamos x y y. Así, si especificamos la posición del planeta en un instante dado, dando x y y (supondremos que z es siempre cero porque no hay fuerza en la dirección z y, si no hay velocidad inicial v_2, no habrá más que hacer z igual a cero), la fuerza está dirigida a lo largo de la línea que une al planeta con el Sol, como se muestra en la figura 9-5.

En esta figura vemos que la componente horizontal de la fuerza está relacionada con la fuerza completa, de la misma manera que la distancia horizontal x lo está a la hipotenusa completa r, porque los dos triángulos son semejantes. También, si x es positivo, F_x es negativo. Esto es, $F_x/|F| = -x/r$, o $F_x = -|F|x/r = -GMmx/r^3$. Ahora usamos la ley dinámica para encontrar que esta componente de la fuerza es igual a la masa del planeta por la variación temporal de su velocidad en la dirección x. Así encontraremos las siguientes leyes:

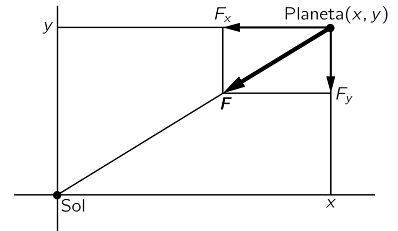


Figura 9-5. La fuerza de gravedad sobre un planeta.

$$m(dv_x/dt) = -GMmx/r^3,$$
$$m(dv_y/dt) = -GMmy/r^3, \qquad (9.17)$$
$$r = \sqrt{x^2 + y^2}.$$

Éste es, entonces, el conjunto de ecuaciones que debemos resolver. Otra vez, para simplificar el trabajo numérico, supondremos que la unidad de tiempo, o la masa del Sol, ha sido ajustada en forma tal (o que la suerte está con nosotros) que $GM \equiv 1$. Para nuestro ejemplo específico supondremos que la posición inicial del planeta está en $x = 0.500$ y $y = 0.000$, y que la velocidad está en un comienzo en la dirección y, y que tiene una magnitud de 1.630. Ahora, ¿cómo hacemos el cálculo? Hacemos otra vez una tabla con columnas para el tiempo, la posición x, la velocidad según x v_x, y la aceleración según x a_x; luego, separadas por una doble línea, tres columnas para la posición, la velocidad y la aceleración en la dirección y. Para obtener las aceleraciones vamos a necesitar la ecuación (9.17); ésta nos dice que la aceleración en la dirección x es $-x/r^3$, y la aceleración en la dirección y es $-y/r^3$, y que r es la raíz cuadrada de $x^2 + y^2$. Así, dado x y y, debemos hacer un pequeño cálculo por separado, tomando la raíz cuadrada de la suma de los cuadrados para encontrar r y luego, para estar preparados para calcular las dos aceleraciones, es útil también evaluar $1/r^3$. Este trabajo puede ser hecho bastante fácilmente usando una tabla de cuadrados, cubos e inversos, entonces necesitamos sólo multiplicar x por $1/r^3$, que hacemos con una regla de cálculo. Así nuestro cálculo procede con los siguientes pasos, usando intervalos de tiempo de $\epsilon = 0.100$: valores iniciales en $t = 0$:

$$x(0) = 0.500 \qquad y(0) = 0.000$$
$$v_x(0) = 0.000 \qquad v_y(0) = +1.630$$

A partir de éstos encontramos:

$$r(0) = 0.500 \qquad 1/r^3(0) = 8.000$$
$$a_x(0) = -4.000 \qquad a_y(0) = 0.000$$

Así podemos calcular las velocidades $v_x(0.05)$ y $v_y(0.05)$:

$$v_x(0.05) = 0.000 - 4.000 \times 0.050 = -0.200;$$
$$v_y(0.05) = 1.630 + 0.000 \times 0.050 = 1.630.$$

Ahora comienza nuestro cálculo principal:

$$\begin{aligned}
x(0.1) &= 0.500 - 0.20 \times 0.1 &&= 0.480\\
y(0.1) &= 0.0 + 1.63 \times 0.1 &&= 0.163\\
r(0.1) &= \sqrt{0.480^2 + 0.163^2} &&= 0.507\\
1/r^3(0.1) &= 7.677\\
a_x(0.1) &= -0.480 \times 7.677 &&= -3.685\\
a_y(0.1) &= -0.163 \times 7.677 &&= -1.250\\
v_x(0.15) &= -0.200 - 3.685 \times 0.1 &&= -0.568\\
v_y(0.15) &= 1.630 - 1.250 \times 0.1 &&= 1.505\\
x(0.2) &= 0.480 - 0.568 \times 0.1 &&= 0.423\\
y(0.2) &= 0.163 + 1.505 \times 0.1 &&= 0.313\\
&\text{etcétera.}
\end{aligned}$$

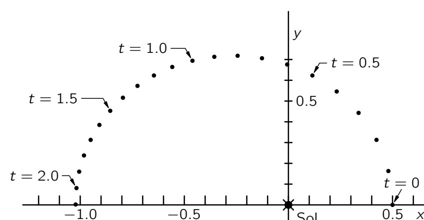


Figura 9-6. El movimiento calculado de un planeta alrededor del Sol.

De esta manera obtenemos los valores dados en la tabla 9-2, ¡y en 20 pasos o algo así hemos seguido la mitad del camino del planeta alrededor del Sol! En la figura 9-6 están dibujadas las coordenadas x y y dadas en la tabla 9-2. Los puntos representan las posiciones en la sucesión de tiempos separados en un décimo de unidad, vemos que al comienzo el planeta se mueve rápidamente y al final se mueve lentamente, y así la forma de la curva queda determinada. ¡Así vemos que *realmente* sabemos cómo calcular el movimiento de los planetas!

Ahora veamos cómo podemos calcular el movimiento de Neptuno, Júpiter, Urano o cualquier otro planeta. Si tenemos un gran número de planetas y hacemos que el Sol se mueva también, ¿podemos hacer la misma cosa? Por supuesto que podemos. Calculamos la fuerza sobre un planeta particular, digamos planeta número i, el cual tiene una posición $x_i; y_i, z_i$ ($i = 1$ puede representar al Sol, $i = 2$ a Mercurio, $i = 3$ a Venus, y así sucesivamente). Debemos conocer las posiciones de todos los planetas. La fuerza que actúa sobre uno se debe a todos los otros cuerpos que están localizados, digamos, en las posiciones x_j, y_j, z_j. Por lo tanto, las ecuaciones son:

$$\begin{aligned}
m_i \frac{dv_{ix}}{dt} &= \sum_{j=1}^{N} -\frac{Gm_i m_j (x_i - x_j)}{r_{ij}^3},\\
m_i \frac{dv_{iy}}{dt} &= \sum_{j=1}^{N} -\frac{Gm_i m_j (y_i - y_j)}{r_{ij}^3},\\
m_i \frac{dv_{iz}}{dt} &= \sum_{j=1}^{N} -\frac{Gm_i m_j (z_i - z_j)}{r_{ij}^3}.
\end{aligned} \tag{9.18}$$

Además, definimos a r_{ij} como la distancia entre los dos planetas i y j; esto es igual a:

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \tag{9.19}$$

También, $\sum$ significa una suma de todos los valores de j —todos los otros cuerpos— excepto, por supuesto, para $j = i$—. Así, todo lo que tenemos que hacer es confeccionar más columnas, *muchas* más columnas. Necesitamos nueve columnas para el movimiento de Júpiter, nueve para el movimiento de Saturno, y así sucesivamente. Entonces cuando tenemos todas las posiciones y velocidades

Tabla 9-2

Solución de $dv_x/dt = -x/r^3$, $dv_y/dt = -y/r^3$, $r = \sqrt{x^2 + y^2}$.
Intervalo: $\epsilon = 0.100$
Órbita $v_y = 1.63$ $v_x = 0$ $x = 0.5$ $y = 0$ y $t = 0$

t	x	v_x	a_x	y	v_y	a_y	r	$1/r^3$
0.0	0.500		−4.000	0.000		0.000	0.500	8.000
		−0.200			1.630			
0.1	0.480		−3.685	0.163		−1.251	0.507	7.677
		−0.568			1.505			
0.2	0.423		−2.897	0.313		−2.146	0.527	6.847
		−0.858			1.290			
0.3	0.337		−1.958	0.443		−2.569	0.556	5.805
		−1.054			1.033			
0.4	0.232		−1.112	0.546		−2.617	0.593	4.794
		−1.165			0.772			
0.5	0.115		−0.454	0.623		−2.449	0.634	3.931
		−1.211			0.527			
0.6	−0.006		+0.018	0.676		−2.190	0.676	3.241
		−1.209			0.308			
0.7	−0.127		+0.342	0.706		−1.911	0.718	2.705
		−1.175			0.117			
0.8	−0.244		+0.559	0.718		−1.646	0.758	2.292
		−1.119			−0.048			
0.9	−0.356		+0.702	0.713		−1.408	0.797	1.974
		−1.048			−0.189			
1.0	−0.461		+0.796	0.694		−1.200	0.833	1.728
		−0.969			−0.309			
1.1	−0.558		+0.856	0.664		−1.019	0.867	1.536
		−0.883			−0.411			
1.2	−0.646		+0.895	0.623		−0.862	0.897	1.385
		−0.794			−0.497			
1.3	−0.725		+0.919	0.573		−0.726	0.924	1.267
		−0.702			−0.569			
1.4	−0.795		+0.933	0.516		−0.605	0.948	1.174
		−0.608			−0.630			
1.5	−0.856		+0.942	0.453		−0.498	0.969	1.100
		−0.514			−0.680			
1.6	−0.908		+0.947	0.385		−0.402	0.986	1.043
		−0.420			−0.720			
1.7	−0.950		+0.950	0.313		−0.313	1.000	1.000
		−0.325			−0.751			
1.8	−0.982		+0.952	0.238		−0.230	1.010	0.969
		−0.229			−0.774			
1.9	−1.005		+0.953	0.160		−0.152	1.018	0.949
		−0.134			−0.790			
2.0	−1.018		+0.955	0.081		−0.076	1.022	0.938
		−0.038			−0.797			
2.1	−1.022		+0.957	0.002		−0.002	1.022	0.936
		+0.057			−0.797			
2.2	−1.017		+0.959	−0.078		+0.074	1.020	0.944
					−0.790			
2.3								

Eje x cruzado a los 2.101 segundos, $\therefore$ periodo $= 4.20$ s
$v_x = 0$ a 2.086 s
Valor de x y -1.022, $\therefore$ semieje mayor $= \dfrac{1.022 + 0.500}{2} = 0.761$
$v_y = -0.797$
Tiempo predicho $\pi(0.761)^{3/2} = \pi(0.663) = 2.082$

iniciales podemos calcular todas las aceleraciones de la ecuación (9.18) calculando primero todas las distancias, usando la ecuación (9.19). ¿Cuánto tiempo tomará para hacerlo? Si lo hacen en casa, ¡necesitarán un tiempo muy largo! Pero en los tiempos modernos tenemos máquinas que hacen la aritmética muy rápidamente, una máquina computadora muy buena puede tomar un microsegundo, esto es, una millonésima de un segundo, para hacer una suma. Para hacer una multiplicación demora más, digamos 10 microsegundos. Puede que en un ciclo de cálculo, dependiendo del problema, tengamos 30 multiplicaciones o algo así, por lo que un ciclo demorará 300 microsegundos. Esto significa que podemos hacer 3 000 ciclos de computación por segundo. Para obtener una exactitud, de, digamos, una parte en 1 000 millones, necesitaríamos 4×10^5 ciclos para corresponder a una revolución de un planeta alrededor del Sol. Esto corresponde a un tiempo de computación de 130 segundos o alrededor de dos minutos. ¡Así, pues, lleva sólo dos minutos para seguir a Júpiter alrededor del Sol, con todas las perturbaciones de todos los planetas corregida a una parte en 1 000 millones, por este método! (Resulta que el error varía casi como el cuadrado del intervalo ϵ. Si hacemos el intervalo 1 000 veces más pequeño, es un millón de veces más exacto. Así que hagamos el intervalo 10 000 veces más pequeño.)

Así, pues, como dijimos, empezamos este capítulo sin saber aún cómo calcular el movimiento de una masa en un resorte. Ahora, armados con el tremendo poder de las leyes de Newton, podemos no sólo calcular tales movimientos simples, sino también, dada sólo una máquina para tratar la aritmética, aun los movimientos tremendamente complejos de los planetas, ¡hasta un grado de precisión tan alto como queramos!

10

Conservación del momentum

10-1 La tercera ley de Newton

Sobre la base de la segunda ley del movimiento de Newton, la cual da la relación entre la aceleración de cualquier cuerpo y la fuerza actuando sobre él, cualquier problema en mecánica puede ser resuelto en principio. Por ejemplo, para determinar el movimiento de unas pocas partículas, uno puede usar el método numérico desarrollado en el capítulo anterior. Pero hay buenas razones para hacer un estudio adicional de las leyes de Newton. Primero, hay casos bastante simples de movimiento que pueden ser analizados no solamente por métodos numéricos, sino que también por análisis matemático directo. Por ejemplo, aunque sabemos que la aceleración de un cuerpo que cae es de $10\,\mathrm{m/s^2}$, y a partir de este hecho podríamos calcular el movimiento por métodos numéricos, es mucho más fácil y satisfactorio analizar el movimiento y encontrar la solución general, $s = s_0 + v_0 t + 5t^2$. De la misma manera, aunque podemos calcular las posiciones de un oscilador armónico por métodos numéricos, es también posible demostrar analíticamente que la solución general es una simple función coseno de t, y así es innecesario recurrir a toda esa dificultad aritmética cuando hay una manera simple y más exacta de obtener el resultado. De la misma manera, aunque el movimiento de un cuerpo alrededor del Sol, determinado por la gravitación, puede ser calculado punto por punto por los métodos numéricos del capítulo 9, lo cual muestra la forma general de la órbita, es interesante también obtener la forma exacta, que el análisis revela como una perfecta elipse.

Desgraciadamente existen en realidad muy pocos problemas que pueden ser resueltos exactamente por análisis. En el caso del oscilador armónico, por ejemplo, si la fuerza del resorte no es proporcional al desplazamiento, sino que es algo más complicada, uno debe volver a caer en el método numérico. Más aún, si hay dos cuerpos girando alrededor del Sol, de modo que el número total de cuerpos es tres, entonces el análisis no puede producir una fórmula simple para el movimiento, y en la práctica el problema debe ser resuelto numéricamente. Ése es el famoso problema de los tres cuerpos, que durante tanto tiempo desafió la capacidad humana de análisis, es muy interesante cuánto necesitó la gente para apreciar el hecho de que tal vez el poder del análisis matemático era limitado y que podría ser necesario usar los métodos numéricos. Hoy día, un enorme número de problemas que no pueden ser resueltos analíticamente se resuelven por métodos numéricos, y el viejo problema de los tres cuerpos, que se suponía difícil, se resuelve como una ejercicio de rutina de una manera exactamente igual a la que fue descrita en el capítulo anterior, es decir, haciendo bastante aritmética. Sin embargo, hay también situaciones donde ambos métodos fallan; los problemas simples pueden hacerse por análisis, y los problemas moderadamente difíciles por métodos aritméticos numéricos, pero los problemas muy complicados no pueden resolverse por ninguno de los dos métodos. Un problema complicado es, por ejemplo, el choque de dos automóviles, o aun el movimiento de las moléculas de un gas. Hay innumerables partículas en un milímetro cúbico de gas, y sería ridículo tratar de hacer cálculos con tantas variables (alrededor de 10^{17}, un décimo de un trillón). Cualquier cosa como el movimiento de las moléculas o átomos de un gas o un pedazo de hierro, o el movimiento de las estrellas en un cúmulo globular, en vez de sólo dos o tres planetas girando alrededor del Sol; tales problemas no pueden resolverse directamente, así que tenemos que buscar otros medios.

En las situaciones en las cuales no podemos seguir el detalle, necesitamos saber algunas propiedades generales, es decir, principios o teoremas generales, los

cuales son consecuencias de las leyes de Newton. Uno de éstos es el principio de conservación de la energía, que se argumentó en el capítulo 4. Otro es el principio de la conservación del momentum, el tema de este capítulo. Otra razón para estudiar más la mecánica es que hay ciertos modelos de movimiento que están repetidos en muchas circunstancias diferentes, así que es bueno estudiar estos modelos para una circunstancia particular. Por ejemplo, estudiaremos los choques, diferentes tipos de choques tienen mucho en común. En el flujo de fluidos, no hace mucha diferencia lo que es el fluido, las leyes del flujo son similares. Otros problemas que estudiaremos son las vibraciones y oscilaciones y, en particular, los fenómenos singulares de las ondas mecánicas —sonido, vibraciones de varillas, etcétera—.

En nuestra discusión de las leyes de Newton se explicó que estas leyes son una especie de programa que dice "Ponga atención a las fuerzas", y que Newton nos dijo sólo dos cosas acerca de la naturaleza de las fuerzas. En el caso de la gravitación, nos dio la ley completa de fuerza. En el caso de las fuerzas muy complicadas entre átomos, él no conocía las leyes correctas para las fuerzas; sin embargo, descubrió una regla, una propiedad general de las fuerzas, que está expresada en su tercera ley, y que es el conocimiento total que Newton tenía acerca de la naturaleza de las fuerzas —la ley de la gravitación y este principio, pero no otros detalles—.

Este principio postula que *la acción es igual a la reacción.*

Lo que se quiere decir es algo de este tipo: supongamos que tenemos dos pequeños cuerpos, digamos partículas, y supongamos que la primera ejerce una fuerza sobre la segunda, empujándola con una cierta fuerza. Entonces simultáneamente, de acuerdo a la tercera ley de Newton, la segunda partícula empujará a la primera con una fuerza igual, en dirección opuesta, más aún, estas fuerzas actúan efectivamente en la misma línea. Ésta es la hipótesis o ley que Newton propuso, y parece ser completamente precisa, aunque no es exacta (discutiremos los errores más tarde). Por el momento tomaremos por verdadero que la acción es igual a la reacción. Por supuesto, si hay una tercera partícula, no en la misma línea que las otras dos, la ley *no* significa que la fuerza total sobre la primera sea igual a la fuerza total sobre la segunda, ya que la tercera partícula, por ejemplo, ejerce su propio empuje sobre cada una de las otras dos. El resultado es que el efecto total sobre las primeras dos es en alguna otra dirección, y las fuerzas sobre las primeras dos partículas no son, en general, ni iguales ni opuestas. Sin embargo, las fuerzas sobre cada partícula pueden descomponerse en partes, habiendo una contribución o parte debida a cada otra partícula que está interactuando. Entonces cada *par* de partículas tiene componentes correspondientes de interacción mutua que son iguales en módulo y opuestas en dirección.

10-2 Conservación del momentum

¿Cuáles son ahora las consecuencias interesantes de la relación anterior? Supongamos, por simplicidad, que tenemos sólo dos partículas interactuando, posiblemente de masas diferentes, y numeradas con 1 y 2. Las fuerzas entre ellas son iguales y opuestas, ¿cuáles son las consecuencias? De acuerdo con la segunda ley de Newton, la fuerza es la rapidez de variación del momentum, así concluimos que la rapidez de variación del momentum p_1 de la partícula 1 es igual a menos la rapidez de variación del momentum p_2 de la partícula 2, o bien:

$$dp_1/dt = -dp_2/dt. \tag{10.1}$$

Ahora, si la *rapidez de variación* es siempre igual y opuesta, se concluye que la *variación total* del momentum de la partícula 1 es igual y opuesta a la *variación total* del momentum de la partícula 2; esto significa que si *sumamos* el momentum de la partícula 1 al momentum de la partícula 2, la rapidez de variación de la

suma de las dos, debido a las fuerzas mutuas (llamadas fuerzas internas) entre partículas, es cero, esto es:

$$d(p_1 + p_2)/dt = 0. \tag{10.2}$$

Se ha supuesto que no hay otras fuerzas en el problema. Si la rapidez de variación de esta suma es siempre cero, esto es sólo otra manera de decir que la cantidad $(p_1 + p_2)$ no cambia. (Esta cantidad también se escribe como $m_1v_1 + m_2v_2$, y se llama *momentum total* de las dos partículas.) Hemos obtenido ahora el resultado de que el momentum total de las dos partículas no se modifica a causa de ningún tipo de interacciones mutuas entre ellas. Ésta afirmación habla de la ley de conservación del momentum en este ejemplo particular. Concluimos que si hay cualquier tipo de fuerza, no importa lo complicada que sea, entre dos partículas, y si medimos o calculamos $m_1v_1 + m_2v_2$, esto es, la suma de los dos momentum antes y después de que las fuerzas actúen, los resultados serán iguales, es decir, el momentum total es constante.

Si extendemos el razonamiento a tres o más partículas interactuando en circunstancias más complicadas, es evidente que en lo que concierne a las fuerzas internas, el momentum total de todas las partículas permanece constante, ya que un aumento del momentum de una debido a otra está exactamente compensado por la disminución del segundo debido a la primera. Esto es, todas las fuerzas internas se equilibrarán y por lo tanto el momentum total de las partículas no puede cambiar. Entonces, si no hay fuerzas desde el exterior (fuerzas externas), no hay fuerzas que puedan cambiar el momentum total, de aquí que el momentum sea una constante.

Es importante describir lo que sucede cuando hay fuerzas que *no* provienen de acciones recíprocas de las partículas en cuestión. Supongamos que aislamos las partículas que interactúan. Si sólo hay fuerzas mutuas, entonces, como antes, el momentum de las partículas no cambia, no importa lo complicadas que sean las fuerzas. Por otro lado, supongamos también que hay fuerzas provenientes de partículas fuera del grupo aislado. Cualquier fuerza ejercida por cuerpos externos sobre cuerpos internos la llamamos una fuerza *externa*. Demostraremos más tarde que la suma de todas las fuerzas externas es igual a la rapidez de variación del momentum total de todas las partículas internas, un teorema muy útil.

La conservación del momentum total de un número de partículas que interactúan puede ser expresada como una constante,

$$m_1v_1 + m_2v_2 + m_3v_3 + \cdots = \text{una constante}, \tag{10.3}$$

si no existen fuerzas netas externas. Aquí las masas y las correspondientes velocidades de las partículas se numeran 1, 2, 3, 4... El enunciado general de la segunda ley de Newton para cada partícula es:

$$F = \frac{d}{dt}(mv), \tag{10.4}$$

es verdadero específicamente para las *componentes* de la fuerza y del momentum para cualquier dirección dada, así la componente x de la fuerza sobre una partícula es igual a la componente x de la rapidez de variación del momentum de esa partícula, o bien:

$$F_x = \frac{d}{dt}(mv_x), \tag{10.5}$$

y en forma similar para las direcciones y y z. Por lo tanto, la ecuación (10.3) es realmente tres ecuaciones, una para cada dirección.

Además de la ley de conservación del momentum, hay otra interesante consecuencia de la segunda ley de Newton, que será demostrada más adelante, pero que simplemente establecemos ahora. Este principio es que las leyes de la física serán las mismas, ya sea que estemos quietos o moviéndonos con una velocidad uniforme en línea recta. Por ejemplo, un niño que hace rebotar una pelota en un avión encuentra que la pelota rebota lo mismo que si la estuviera haciendo

rebotar sobre la tierra. Aun cuando el avión se está moviendo con una velocidad muy alta, a menos que cambie su velocidad, las leyes parecen las mismas para el niño que cuando el avión está en reposo. Esto es llamado *principio de relatividad*. Tal como lo usamos aquí lo llamaremos “relatividad galileana”, para distinguirla del análisis más cuidadoso hecho por Einstein, que estudiaremos más tarde.

Hemos ya deducido la ley de conservación del momentum a partir de las leyes de Newton y podríamos continuar desde aquí para encontrar las leyes especiales que describen los impactos y colisiones. Pero por razones de variedad, y también como una ilustración de un tipo de razonamiento que puede ser usado en física en otras circunstancias, donde, por ejemplo, uno no puede conocer las leyes de Newton y pueda seguir un método diferente, discutiremos las leyes de los impactos y colisiones desde un punto de vista completamente diferente. Basaremos nuestra discusión en el principio de la relatividad galileana, enunciado antes, y terminaremos con la ley de conservación del momentum.

Comenzaremos suponiendo que si corremos, la naturaleza se verá igual que si estamos quietos. Antes de discutir las colisiones en las cuales dos cuerpos chocan y se pegan, o se juntan y rebotan apartándose, consideraremos primero dos cuerpos que se mantienen unidos por un resorte o algo por el estilo, y luego se suelta repentinamente y el resorte o tal vez una pequeña explosión los empuja. Más aún, consideraremos el movimiento en una sola dirección. Supongamos primero que los dos objetos son exactamente iguales, son objetos bien simétricos, y luego tenemos una pequeña explosión entre ellos. Después de la explosión, uno de los cuerpos se estará moviendo, digamos hacia la derecha, con una velocidad v. Entonces parece razonable que el otro cuerpo se esté moviendo hacia la izquierda con una velocidad v, porque si los objetos son semejantes, no hay razón para que la derecha o la izquierda sean preferidas y, por lo tanto, los cuerpos harían algo que es simétrico. Esto es un ejemplo de un tipo de ideas que es muy útil en muchos problemas, pero no se habría puesto al descubierto si hubiéramos comenzado con las fórmulas.

El primer resultado de nuestro experimento es que objetos iguales tendrán velocidades iguales, pero ahora supongamos que tenemos dos objetos hechos de diferentes materiales, digamos cobre y aluminio, y hacemos las dos *masas* iguales. Ahora supondremos que si hacemos el experimento con dos masas que son iguales, aun cuando los objetos no son idénticos, las velocidades serán iguales. Alguien podría objetar: “Pero, claro, podría hacerlo a la inversa, no tenía por qué *suponer* eso. Usted podría *definir* que masas iguales significa dos masas que adquieren velocidades iguales en este experimento”. Sigamos esta sugerencia y hagamos una pequeña explosión entre el cobre y un pedazo muy grande de aluminio, tan pesado que el cobre vuela lejos y el aluminio apenas se mueve. Eso es demasiado aluminio, así que reduzcamos la cantidad hasta que haya sólo un pedazo muy diminuto; entonces, cuando hacemos la explosión, el aluminio vuela lejos y el *cobre* apenas se mueve. Esto no es suficiente aluminio. Evidentemente hay alguna cantidad correcta intermedia, así que seguimos ajustando las cantidades hasta que las velocidades lleguen a ser iguales. Muy bien; entonces, demos vuelta a las cosas y digamos que cuando las velocidades son iguales, las masas son iguales. Esto parece ser una definición, y parece notable que podamos transformar las leyes físicas en meras definiciones. Sin embargo, *hay* algunas leyes físicas en juego, y si aceptamos esta descripción de masas iguales, inmediatamente encontramos una de las leyes, como mencionaremos.

Supongamos que sabemos, a partir del experimento anterior, que dos pedazos de materia A y B (de cobre y de aluminio) tienen masas iguales y comparamos un tercer cuerpo, digamos un pedazo de oro, con el cobre de la misma manera anterior, asegurándonos de que su masa sea igual a la masa del cobre. Si ahora hacemos el experimento entre el aluminio y el oro, no hay nada lógico que diga que *estas* masas deban ser iguales. Sin embargo, el *experimento* demuestra que realmente lo son. Así ahora, por el experimento, hemos encontrado una nueva ley. Un enunciado de esta ley puede ser: Si dos masas son iguales a una tercera masa (como se determina por las velocidades iguales en este experimento), entonces son iguales entre sí. (Este enunciado *no* es en absoluto consecuencia de un enunciado

similar usado como postulado en relación con cantidades *matemáticas*.) A partir de este ejemplo podemos ver lo rápido que comenzamos a inferir cosas si somos descuidados. *No* es precisamente una definición decir que las masas son iguales cuando las velocidades son iguales, porque decir que las masas son iguales implica las leyes matemáticas de la igualdad, lo que a su vez hace una predicción acerca de un experimento.

Como segundo ejemplo, supongamos que se encuentra que A y B son iguales al hacer el experimento con una cierta intensidad de explosión, que da una cierta velocidad; si después usamos una explosión más fuerte, ¿será verdad o no con respecto a las velocidades obtenidas ahora sean iguales? Otra vez, en lógica nada hay que pueda decidir esta cuestión, pero el experimento demuestra que *es* verdadero. Así, hay otra ley que se puede establecer: si dos cuerpos tienen masas iguales cuando se las mide por velocidades iguales a una velocidad determinada, tendrán masas iguales cuando se las mide a otra velocidad. A partir de estos ejemplos, vemos que lo que pareció ser sólo una definición realmente encierra algunas leyes de física.

En el desarrollo que sigue supondremos que es verdad que masas iguales tienen velocidades iguales y opuestas cuando ocurre una explosión entre ellas. Haremos otra suposición en el caso inverso: si dos objetos idénticos, moviéndose en direcciones opuestas con velocidades iguales, chocan y se pegan por algún tipo de cola, entonces ¿de qué manera se estarán moviendo después de la colisión? Ésta es otra vez una situación de simetría, sin preferencia entre derecha e izquierda; así que suponemos que quedan en reposo. También supondremos que dos objetos cualesquiera de masas iguales, aun si los objetos están hechos de materiales diferentes, que chocan y se pegan cuando se mueven con la misma velocidad en direcciones opuestas, quedarán en reposo después de la colisión.

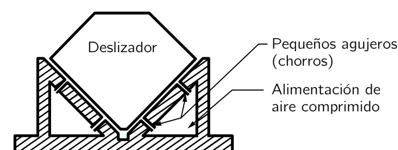


Figura 10-1. Vista desde un extremo de un canal de aire lineal.

10-3 ¡El momentum *se* conserva!

Podemos verificar experimentalmente las hipótesis anteriores. Primero, si dos objetos en reposo con masas iguales son separados por una explosión, se alejarán con la misma velocidad, y segundo, si estos dos objetos de masas iguales se acercan con la misma velocidad, chocan y se pegan, se detendrán. Esto podemos hacerlo con un maravilloso invento llamado canal de aire,* que elimina el roce, la cosa que continuamente molestó a Galileo (figura 10-1). Él no pudo hacer experimentos deslizando las cosas porque no se deslizaban libremente, pero con un toque mágico adicional, hoy podemos obtener la eliminación del roce. Nuestros objetos se deslizarán sin dificultad, continuamente a una velocidad constante, como fue anunciado por Galileo. Esto se hace sosteniendo los objetos sobre aire. Debido a que el aire tiene un roce muy bajo, un objeto se desliza con velocidad prácticamente constante cuando no hay fuerza aplicada. Primero usamos dos bloques deslizantes que se han construido cuidadosamente para que tengan el mismo peso o masa (realmente se midieron sus pesos, pero sabemos que este peso es proporcional a la masa), y colocamos una pequeña cápsula explosiva en un cilindro cerrado entre los dos bloques (figura 10-2). Haremos partir los bloques desde el reposo en el centro de la pista y los obligaremos a apartarse haciendo explotar la cápsula con una chispa eléctrica. ¿Qué sucederá? Si las velocidades son iguales cuando se apartan, llegarán a los extremos del canal al mismo tiempo. Al llegar a los extremos rebotarán prácticamente con velocidades opuestas, y se juntarán y detendrán en el centro desde donde partieron. Es una buena comprobación, cuando se realiza realmente; el resultado es justamente como lo hemos descrito (figura 10-3).

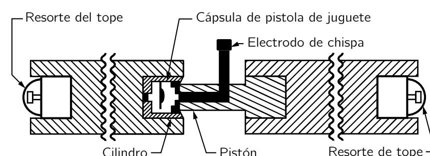


Figura 10-2. Corte de los deslizadores con cilindro accesorio de interacción explosiva.

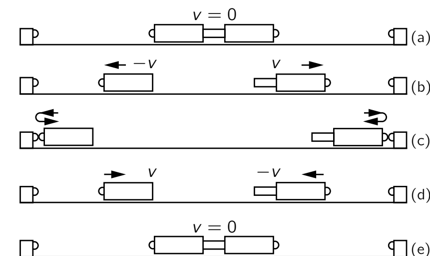


Figura 10-3. Vista esquemática del experimento acción y reacción con masas iguales.

Ahora, la próxima cosa que nos gustaría resolver es qué sucede en una situación menos simple. Supongamos que tenemos dos masas iguales, una moviéndose con velocidad v y la otra en reposo, y que chocan y se pegan, ¿qué sucederá? Hay una masa $2m$ en total cuando hemos terminado, moviéndose con una velocidad desconocida. ¿Qué velocidad? Éste es el problema. Para encontrar la respuesta,

* H. V. Neher y R. B. Leighton, *Amer. Jour. of Phys.* **31**, 255 (1963).

hacemos la suposición de que si viajamos en un coche, físicamente observaremos lo mismo que si estuviéramos en reposo. Partimos sabiendo que dos masas iguales, que se mueven en direcciones opuestas con velocidades iguales v, se pararán en seco cuando choquen. Supongamos ahora que si esto sucede, pasamos en un automóvil, a una velocidad $-v$. ¿Cómo se ve entonces? Puesto que estamos viajando junto a una de las dos masas que se acercan entre sí, nos parece que ésta tiene velocidad cero. La otra masa, sin embargo, moviéndose en el otro sentido con velocidad v, parecerá estar viniendo hacia nosotros con velocidad $2v$ (figura 10-4). Finalmente, las masas combinadas después de la colisión nos parecerán estar pasando con una velocidad v. Por lo tanto, concluiremos que una cosa con velocidad $2v$, chocando con otro igual en reposo, terminará con una velocidad v, o lo que es precisamente igual matemáticamente, un objeto con velocidad v chocando con y pegándose a uno en reposo, hará que un objeto se mueva con velocidad $v/2$. Notemos que si multiplicamos de antemano la masa y la velocidad y las sumamos $mv + 0$, obtenemos el mismo resultado que cuando multiplicamos después la masa y la velocidad del conjunto, $2m$ por $v/2$. Así, esto nos dice lo que sucede cuando una masa de velocidad v choca con una que está en reposo.

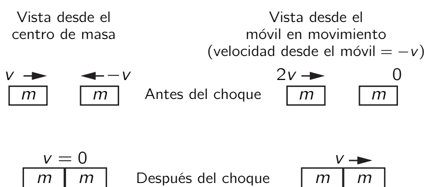


Figura 10-4. Dos vistas de un choque inelástico entre masas iguales.

Exactamente de la misma manera podemos deducir lo que sucede cuando objetos iguales que tienen dos velocidades *cualesquiera* chocan entre sí.

Supongamos que tenemos dos cuerpos iguales con velocidades v_1 y v_2 respectivamente, que chocan y se pegan. ¿Cuál es su velocidad v, después de la colisión? Otra vez pasamos en un automóvil, digamos a una velocidad v_2, de manera tal que uno de los cuerpos parece estar en reposo. Entonces el otro parece tener una velocidad $v_1 - v_2$, y tenemos el mismo caso que teníamos antes. Cuando todo esté terminado, se estarán moviendo con una velocidad $\frac{1}{2}(v_1 - v_2)$ respecto al coche. ¿Cuál es entonces la velocidad real respecto al suelo? Ésta es $v = \frac{1}{2}(v_1 - v_2) + v_2$ o $\frac{1}{2}(v_1 - v_2)$ (figura 10-5). Otra vez notamos que:

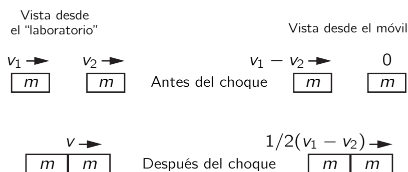


Figura 10-5. Dos vistas de otro choque inelástico entre masas iguales.

$$mv_1 + mv_2 = 2m(v_1 + v_2)/2. \tag{10.6}$$

Así, usando este principio, podemos analizar cualquier tipo de colisión en la cual dos cuerpos de masas iguales chocan entre sí y se pegan. En realidad, aunque hemos trabajado sólo en una dirección, podemos descubrir mucho acerca de colisiones mucho más complicadas, imaginando que pasamos en un coche en alguna dirección oblicua. El principio es el mismo, pero los detalles se hacen algo complicados.

Para verificar experimentalmente si un objeto que se mueve con velocidad v, y que choca con uno igual en reposo, forma un objeto que se mueve con velocidad $v/2$, podemos realizar el siguiente experimento con nuestro aparato del canal de aire. Colocamos en el canal tres objetos, de igual masa, dos de los cuales están inicialmente unidos con nuestro dispositivo cilíndrico explosivo, estando el tercero muy cerca pero ligeramente separado de aquéllos y provisto de un tope adhesivo de forma que se pegará con otro objeto que lo choque. Así, un momento después de la explosión, tenemos dos objetos de masa m moviéndose con velocidades v iguales y opuestas. Un momento después de esto, uno de éstos choca con el tercer objeto y hace moverse un objeto de masa $2m$, creemos así, con velocidad $v/2$. ¿Cómo podemos verificar que es realmente $v/2$? Arreglando las posiciones iniciales de las masas sobre el canal, de manera que las distancias a los extremos no sean iguales, pero que estén en la razón 2 : 1. Así, nuestra primera masa, la cual continúa moviéndose con velocidad v, recorrerá en un tiempo dado el doble de distancia que las dos que se han pegado (prescindiendo de la pequeña distancia recorrida por el segundo objeto antes que choque con el tercero). La masa m y la masa $2m$ alcanzarán los extremos al mismo tiempo y, cuando lo experimentamos, encontramos que lo hacen (figura 10-6).

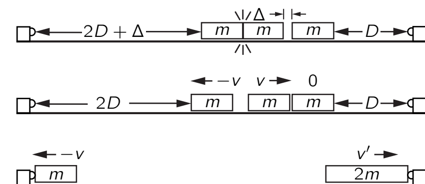


Figura 10-6. Una experiencia para verificar que una masa m con velocidad v que choca contra una masa m con velocidad cero da como resultado $2m$ con velocidad $v/2$.

El próximo problema que queremos desarrollar es qué sucede si tenemos dos masas diferentes. Tomemos una masa m y una masa $2m$ y apliquemos nuestra interacción explosiva. ¿Qué sucederá entonces? Si como resultado de la explosión, m se mueve con velocidad v, ¿con qué velocidad se mueve $2m$? El experimento que recién hemos realizado se puede repetir con separación nula entre la segunda

y tercera masa, y cuando lo ensayamos obtenemos el mismo resultado, es decir, las masas que reaccionan m y $2m$ alcanzan las velocidades $-v$ y $v/2$. Así, la reacción directa entre m y $2m$ da el mismo resultado que la reacción simétrica entre m y m, seguida de una colisión entre m y una tercera masa m a la cual se pega. Más aún, encontramos que las masas m y $2m$, al volver desde los extremos del canal con sus velocidades (casi) exactamente invertidas, se paran en seco si se pegan.

Ahora la próxima cuestión que podemos preguntarnos es ésta: ¿qué sucederá si una masa m con velocidad v, digamos, choca y se pega a otra masa $2m$ en reposo? Esto es muy fácil de responder usando el principio de relatividad galileana, porque simplemente observamos la colisión que recién hemos descrito desde un automóvil moviéndose con velocidad $-v/2$ (figura 10-7). Desde el automóvil, las velocidades son:

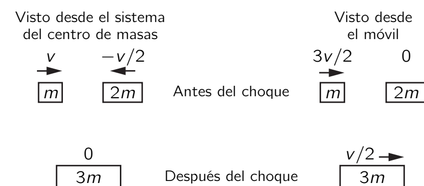


Figura 10-7. Dos vistas de un choque inelástico entre m y $2m$.

$$v_1' = v - v(\text{auto}) = v + v/2 = 3v/2$$

y

$$v_2' = -v/2 - v(\text{auto}) = -v/2 + v/2 = 0.$$

Después de la colisión la masa $3m$ nos parece moverse con velocidad $v/2$. Así obtenemos la respuesta, es decir, la razón de las velocidades antes y después de la colisión es 3 a 1. Si un objeto de masa m choca con un objeto en reposo de masa $2m$, el conjunto se aleja, pegado, con 1/3 de velocidad. La regla general otra vez es que la suma de los productos de las masas y las velocidades permanece igual: $mv + 0$ igual a $3m$ por $v/3$, por lo que estamos construyendo gradualmente el teorema de conservación del momentum parte por parte.

Ahora tenemos uno contra dos. Usando los mismos razonamientos podemos predecir los resultados de uno contra tres, dos contra tres, etc. El caso de dos contra tres, comenzando desde el reposo, se muestra en la figura 10-8.

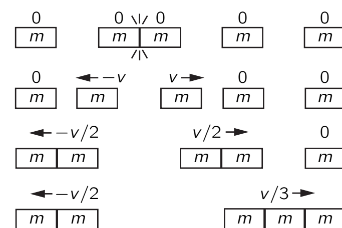


Figura 10-8. Acción y reacción entre $2m$ y $3m$.

En cada caso encontramos que la masa del primer objeto por su velocidad, más la masa del segundo objeto por su velocidad, es igual a la masa total del objeto final por su velocidad. Éstos son ejemplos de la conservación del momentum. Partiendo de casos simétricos simples, hemos demostrado la ley para casos más complejos. Podríamos, en realidad, hacerlo para cualquier fracción racional de masa, y como cada fracción está próxima a una fracción racional, podemos manejar cualquier razón tanto como queramos.

10-4 Momentum y energía

Todos los ejemplos anteriores son casos simples donde los cuerpos chocan y se quedan unidos o estaban inicialmente unidos y posteriormente eran separados por una explosión. Sin embargo, hay situaciones en las cuales los cuerpos *no* se unen como, por ejemplo, dos cuerpos de masas iguales que chocan con velocidades iguales y después rebotan. Por un breve instante están en contacto y ambos están comprimidos. En el instante de máxima compresión ambos tienen velocidad cero y se almacena energía en los cuerpos elásticos, como en un resorte comprimido. Esta energía proviene de la energía cinética que tenían los cuerpos antes de la colisión, la cual llega a ser cero en el instante en que su velocidad es cero. Sin embargo, la pérdida de energía cinética es sólo momentánea. La condición en compresión es análoga a la cápsula que libera energía en una explosión. Los cuerpos son inmediatamente descomprimidos en una especie de explosión y se separan de nuevo, pero ya conocemos el caso —los cuerpos se separan con velocidades iguales—. Sin embargo, esta velocidad de rebote es menor, en general, que la velocidad inicial, porque no toda energía está disponible para la explosión, dependiendo del material. Si el material es como la masilla, la energía cinética no se recobra, pero si es algo más rígido, algo de energía cinética se recupera en general. En la colisión el resto de la energía cinética se transforma en calor y energía de vibración —los cuerpos están calientes y vibrando—. La energía de vibración también se transforma pronto en calor. Es posible hacer los cuerpos

que chocan de materiales altamente elásticos, como el acero con topes de resorte cuidadosamente diseñados, de tal modo que la colisión genera muy poco calor y vibración. En estas circunstancias las velocidades de rebote son prácticamente iguales a las velocidades iniciales; tal colisión se llama *elástica*.

Que la rapidez *antes* y *después* de una colisión elástica sea igual, no es cuestión de conservación del momentum, sino de conservación de la *energía cinética*. Que las velocidades de los cuerpos que rebotan después de una colisión simétrica sean iguales *entre* sí es, sin embargo, una cuestión de conservación del momentum.

Análogamente podemos analizar colisiones entre cuerpos de diferentes masas, de distintas velocidades iniciales, de varios grados de elasticidad, y determinar las velocidades finales y la pérdida de energía cinética, pero no iremos a los detalles de estos procesos.

Las colisiones elásticas son especialmente interesantes para sistemas que no tengan "engranajes, ruedas o partes internas". Entonces cuando hay una colisión no hay ningún lugar donde la energía pueda ser encerrada, porque los objetos que se separan están en la misma condición de que cuando chocaron. Por lo tanto, entre objetos muy elementales, las colisiones son siempre elásticas o muy cercanamente de ser elásticas. Por ejemplo, se dice que las colisiones entre átomos o moléculas en un gas son *perfectamente* elásticas. Aunque esto es una excelente aproximación, hasta esas colisiones no son perfectamente elásticas, de otra manera uno no podría entender cómo la energía en forma de luz o de radiación térmica pueda salir de un gas. De vez en cuando, en una colisión de gases se emite un rayo infrarrojo de baja energía, pero este hecho es muy raro y la energía emitida es muy pequeña. Así, la mayoría de las veces, las colisiones de moléculas en gases se consideran perfectamente elásticas.

Como ejemplo interesante consideremos una colisión *elástica* entre dos objetos de *masas iguales*. Si se acercan con la misma velocidad, se separarán con la misma velocidad por simetría. Pero ahora observemos esto en otra circunstancia, en la cual uno de ellos se está moviendo con velocidad v y el otro está en reposo. ¿Qué sucede? Hemos pasado por esto antes. Observamos la colisión simétrica desde un automóvil que se está moviendo junto con uno de los objetos, y encontramos que si un cuerpo choca elásticamente con otro cuerpo en reposo de exactamente la misma masa, el cuerpo en movimiento se detiene, y el que estaba en reposo se mueve ahora con la misma velocidad que tenía el otro, los cuerpos simplemente intercambian las velocidades. Esta conducta puede ser demostrada fácilmente con un aparato adecuado para el estudio de impactos. En forma más general, si ambos cuerpos se están moviendo con velocidades diferentes, simplemente intercambian velocidad en el impacto.

Otro ejemplo de una interacción casi elástica es el magnetismo. Si disponemos de un par de imanes en forma de U en nuestros bloques deslizantes tal que se repelan, cuando uno se mueve suavemente hacia el otro, lo empuja y queda perfectamente quieto y ahora el otro se aleja moviéndose sin fricción.

El principio de conservación del momentum es muy útil, porque nos permite resolver muchos problemas sin conocer los detalles. No conocíamos los detalles de los movimientos del gas en la explosión de la cápsula; sin embargo, pudimos predecir, por ejemplo, las velocidades con que los cuerpos se separaban. Otro ejemplo interesante es el cohete a propulsión. Un cohete de gran masa, M, expulsa una pequeña porción de masa m, con una enorme velocidad V relativa al cohete. Después de esto el cohete, si estaba originalmente en reposo, se moverá con una pequeña velocidad, v. Usando el principio de conservación del momentum podemos calcular que esta velocidad será:

$$v = \frac{m}{M} \cdot V.$$

Mientras se esté expulsando material, el cohete continúa aumentando su velocidad. La propulsión de cohetes es esencialmente lo mismo que el retroceso de un cañón, no hay necesidad de ningún aire contra el cual empujar.

10-5 Momentum relativista

En los tiempos modernos, la ley de conservación del momentum ha sufrido ciertas modificaciones. Sin embargo, aún hoy la ley es cierta, y sólo se modifica en las definiciones. En la teoría de la relatividad resulta que tenemos conservación del momentum, las partículas tienen masas y el momentum aún está dado por mv, la masa por la velocidad, *pero la masa cambia con la velocidad.* Por lo tanto, el momentum también cambia. La masa varía con la velocidad de acuerdo a la ley:

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \tag{10.7}$$

donde m_0 es la masa del cuerpo en reposo y c es la velocidad de la luz. Es fácil ver en la fórmula que hay una diferencia despreciable entre m y m_0 a menos que v sea muy grande, y que para las velocidades ordinarias la expresión del momentum se reduce a la antigua fórmula.

Las componentes del momentum para una partícula se escriben en la forma:

$$p_x = \frac{m_0 v_x}{\sqrt{1 - v^2/c^2}}, \qquad p_y = \frac{m_0 v_y}{\sqrt{1 - v^2/c^2}}, \qquad p_z = \frac{m_0 v_z}{\sqrt{1 - v^2/c^2}}, \tag{10.8}$$

donde $v^2 = v_x^2 + v_y^2 + v_z^2$. Si las componentes x se suman sobre todas las partículas que interactúan tanto antes como después de una colisión, las sumas son iguales, esto es, el momentum se conserva en la dirección x. Lo mismo vale para cualquier dirección.

En el capítulo 4 vimos que la ley de la conservación de la energía no es válida a menos que reconozcamos que la energía aparece en formas diferentes, energía eléctrica, energía mecánica, energía radiante, energía térmica, y así sucesivamente. En algunos de estos casos, por ejemplo la energía térmica, se puede decir que la energía está "escondida". Este ejemplo puede sugerir la pregunta: "¿Hay también formas de momentum escondidas —tal vez momentum calórico—?" La respuesta es que es muy difícil esconder momentum por las razones siguientes.

Los movimientos al azar de los átomos de un cuerpo dan una medida de la energía térmica si se suman los *cuadrados* de las velocidades. Esta suma será un resultado positivo, no teniendo un carácter direccional. El calor está allí, ya sea que el cuerpo se mueva o no como un todo, y la conservación de la energía en la forma de calor no es muy evidente. Por otro lado, si uno suma las *velocidades*, que tienen dirección, y encuentra un resultado que no es cero, esto significa que existe un cierto desplazamiento del cuerpo entero en alguna dirección particular, y un momentum tan grande se observa fácilmente. Así no hay pérdida al azar de momentum interno, debido a que el cuerpo tiene momentum neto sólo cuando se mueve como un todo. Por lo tanto el momentum, como cantidad mecánica, es difícil de esconder. Sin embargo, el momentum *puede* estar escondido —en el campo electromagnético, por ejemplo—. Este caso es otro efecto de la relatividad.

Una de las proposiciones de Newton fue que las interacciones a distancia son instantáneas. Resulta que éste no es el caso, en situaciones en que intervienen fuerzas eléctricas, por ejemplo, si una carga eléctrica en un lugar se mueve repentinamente, el efecto sobre otra carga, en otro lugar, no aparece instantáneamente —hay un pequeño retardo—. En esas circunstancias, aun si las fuerzas son iguales, las cosas no andan bien con el momentum, habrá un corto tiempo durante el cual existirá dificultad porque por un rato la primera carga sentirá una cierta fuerza de reacción, digamos, y tomará algún momentum, pero la segunda carga no ha sentido nada y no ha cambiado aún su cantidad de movimiento. Lleva tiempo para la influencia cruzar la distancia que interviene, lo que hace a 300 000 kilómetros por segundo. En ese pequeñísimo tiempo el momentum de las partículas no se conserva. Por supuesto, después que la segunda carga ha sentido el efecto de la primera y todo se ha aquietado, se verificará la ecuación del momentum, pero durante ese pequeño intervalo el momentum no se conserva. Representamos esto diciendo que durante este intervalo hay otro tipo de momentum además

del de la partícula, mv, y ése es el momentum del campo electromagnético. Si sumamos el momentum del campo al momentum de las partículas, el momentum se conserva todo el tiempo y en cualquier instante. El hecho de que el campo electromagnético pueda poseer momentum y energía hace ese campo muy real, y así, para una mejor comprensión, la idea original de que hay sólo las fuerzas entre partículas se debe transformar en la idea de que una partícula crea un campo, y un campo actúa sobre otra partícula y el campo mismo tiene propiedades tan familiares como contenido de energía y momentum, como las partículas. Para tomar otro ejemplo, un campo electromagnético tiene ondas, que llamamos luz; resulta que la luz también transporta momentum, de modo que cuando la luz incide sobre un objeto le entrega cierto momentum por segundo, esto es equivalente a una fuerza, porque si el objeto iluminado está recogiendo una cierta porción de momentum por segundo, su cantidad de movimiento está cambiando y la situación es exactamente igual que si hubiera una fuerza sobre él. La luz puede ejercer presión al bombear un objeto, esta presión es muy pequeña, pero es medible con aparatos suficientemente delicados.

Ahora bien, en mecánica cuántica resulta que el momentum es una cosa diferente —ya no es más mv—. Es difícil definir exactamente lo que se quiere decir por velocidad de una partícula, pero el momentum aún existe. En mecánica cuántica la diferencia es que cuando las partículas son representadas como partículas, el momentum es aún mv, pero cuando las partículas son representadas como ondas, el momentum se mide por el número de ondas por centímetro; mientras más grande es este número de ondas más grande es el momentum. A pesar de las diferencias, la ley de conservación del momentum también es válida en la mecánica cuántica. ¡Aun cuando la ley $F = ma$ es falsa, y todas las deducciones de Newton para la conservación del momentum estaban erradas, en la mecánica cuántica, sin embargo, en última instancia, esa ley particular se mantiene!

11

Vectores

11-1 Simetría en física

En este capítulo introducimos un tema que se conoce técnicamente en física como *simetría en las leyes físicas*. La palabra "simetría" se usa aquí con un significado especial, y por esta razón necesita ser definida. ¿Cuándo es simétrica una cosa? —¿cómo podemos definirlo?—. Cuando tenemos un cuadro que es simétrico, un lado es de algún modo igual al otro lado. El profesor Hermann Weyl ha dado esta definición de simetría: una cosa es simétrica si uno puede someterla a una cierta operación y aparece exactamente igual después de la operación. Por ejemplo, si observamos la silueta de un jarrón que es simétrico, y luego lo giramos 180° alrededor del eje vertical, entonces se verá igual. Adoptaremos la definición de simetría de Weyl en una forma más general y de esa manera discutiremos la simetría de las leyes físicas.

Supongan que construimos una máquina compleja en algún lugar, con muchas interacciones complicadas y bolas rebotando por doquier con fuerzas entre sí, etc. Supongan ahora que construimos exactamente la misma clase de equipo en algún otro lugar igualando parte por parte, con las mismas dimensiones y la misma orientación, todo igual sólo que desplazado lateralmente en cierta distancia. Entonces, si hacemos funcionar las dos máquinas en las mismas circunstancias iniciales en exacta correspondencia, nos preguntamos: ¿Se comportará una máquina exactamente igual a la otra? ¿Seguirán todos los movimientos en exacto paralelismo? Por supuesto, la respuesta bien puede ser *no*, porque si escogemos un lugar equivocado para nuestra máquina, ésta podría estar rodeada por una pared, e interferencias desde la pared harían que la máquina no trabajase.

Todas nuestras ideas en física requieren una cierta cantidad de sentido común en sus aplicaciones; no son ideas puramente matemáticas o abstractas. Tenemos que comprender qué queremos decir cuando decimos que los fenómenos son los mismos cuando movemos el aparato a una nueva posición. Queremos decir que movemos todo lo que creemos pertinente; si el fenómeno no es el mismo, sugerimos que algo pertinente no ha sido movido y procedemos a buscarlo. Si nunca lo encontramos, entonces afirmamos que las leyes de la física no tienen esta simetría. Por otro lado, podríamos encontrarlo —esperamos encontrarlo— si las leyes de la física tienen esa simetría; observando alrededor, podemos descubrir, por ejemplo, que la pared está empujando el aparato. La cuestión básica es, si definimos las cosas lo suficientemente bien, si todas las fuerzas esenciales están incluidas dentro del aparato, si todas las partes pertinentes se mueven desde un lugar a otro, ¿serán las leyes las mismas? ¿La maquinaria trabajará de la misma manera?

Está claro que lo que queremos hacer es mover todo el equipo e influencias *esenciales*, pero no *todas las cosas* del mundo —planetas, estrellas y todo— porque si hacemos eso, tenemos de nuevo el mismo fenómeno por la razón trivial de que hemos vuelto atrás exactamente donde comenzamos. No, no podemos moverlo *todo*. Pero resulta en la práctica que con una cierta cantidad de inteligencia acerca de qué mover, la maquinaria trabajará. En otras palabras, si no estamos rodeados por una pared, si conocemos el origen de las fuerzas externas y disponemos que ésas se muevan también, entonces la maquinaria *trabajará* igual en un lugar como en otro.

11-2 Traslaciones

Limitaremos nuestro análisis sólo a la mecánica, para lo cual tenemos ahora suficiente conocimiento. En los capítulos anteriores hemos visto que las leyes de la mecánica pueden ser resumidas en un conjunto de tres ecuaciones para cada partícula:

$$m(d^2x/dt^2) = F_x, \qquad m(d^2y/dt^2) = F_y, \qquad m(d^2z/dt^2) = F_z. \tag{11.1}$$

Ahora bien, esto significa que existe una manera de *medir* x, y y z sobre tres ejes perpendiculares y las fuerzas según esas direcciones, de modo que estas leyes sean válidas. Éstas deben ser medidas desde algún origen, pero, *¿dónde ponemos el origen?* Todo lo que Newton nos diría al principio es que *hay* algún lugar desde el cual podemos medir, quizás el centro del universo, tal que estas leyes sean correctas. Pero podemos demostrar inmediatamente que nunca podemos encontrar el centro, porque si usamos algún otro origen no habría diferencia. En otras palabras, supongamos que hay dos personas: Juan, quien tiene su origen en un lugar, y Pedro, quien tiene un sistema paralelo cuyo origen está en alguna otra parte (figura 11-1). Ahora bien, cuando Juan mide la ubicación del punto en el espacio, lo encuentra en x, y, z (usualmente dejaremos a z fuera porque introduce demasiada confusión al dibujarla en una figura). Pedro, por otro lado, cuando mida el mismo punto, obtendrá un x diferente (para distinguirlo lo llamaremos x'), y en principio un y diferente, aun cuando en nuestro ejemplo son numéricamente iguales. Así, se tiene

$$x' = x - a, \qquad y' = y, \qquad z' = z. \tag{11.2}$$

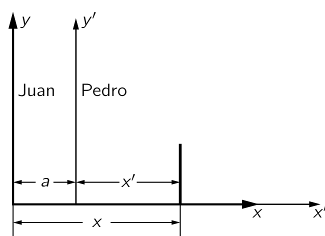


Figura 11-1. Dos sistemas de coordenadas paralelas.

Ahora, a fin de completar nuestro análisis debemos saber lo que Pedro obtendría para las fuerzas. La fuerza se supone que actúa a lo largo de alguna línea y por la fuerza en la dirección x entenderemos la parte del total que está en la dirección x, la cual es el módulo de la fuerza por el coseno de su ángulo con el eje x. Ahora vemos que Pedro usaría exactamente la misma proyección que usaría Juan, así que tenemos un conjunto de ecuaciones

$$F_{x'} = F_x, \qquad F_{y'} = F_y, \qquad F_{z'} = F_z. \tag{11.3}$$

Éstas serían las relaciones entre las cantidades vistas por Juan y Pedro.

La pregunta es, si Juan conoce las leyes de Newton y si Pedro trata de escribir las leyes de Newton, ¿serán también correctas para él? ¿Tiene alguna importancia desde qué origen medimos los puntos? En otras palabras, suponiendo que las ecuaciones (11.1) son verdaderas y que las ecuaciones (11.2) y (11.3) dan la relación entre las medidas, ¿es o no verdad que

$$\begin{aligned} &\text{(a)} \quad m(d^2x'/dt^2) = F_{x'}, \\ &\text{(b)} \quad m(d^2y'/dt^2) = F_{y'}, \\ &\text{(c)} \quad m(d^2z'/dt^2) = F_{z'}? \end{aligned} \tag{11.4}$$

Para verificar estas ecuaciones derivaremos la fórmula para x' dos veces. Primero que nada

$$\frac{dx'}{dt} = \frac{d}{dt}(x - a) = \frac{dx}{dt} - \frac{da}{dt}.$$

Ahora supondremos que el origen de Pedro está fijo (sin moverse) respecto al de Juan; por lo tanto, a es una constante y $da/dt = 0$, así encontramos que

$$dx'/dt = dx/dt$$

y por lo tanto

$$d^2x'/dt^2 = d^2x/dt^2;$$

sabemos, pues, que la ecuación (11.4a) se vuelve

$$m(d^2x/dt^2) = F_{x'}.$$

(También suponemos que las masas medidas por Juan y Pedro son iguales.) Así, la aceleración por la masa es la misma que la del otro. Hemos encontrado también la fórmula para $F_{x'}$, puesto que sustituyendo de la ecuación (11.1) encontramos que

$$F_{x'} = F_x.$$

Por lo tanto, las leyes tal como las ve Pedro aparecen iguales; él también puede escribir las leyes de Newton, con diferentes coordenadas, y siguen siendo correctas. Eso significa que no hay una manera única de definir el origen del mundo, porque las leyes tienen la misma forma, en cualquier posición desde la cual son observadas.

Esto es también verdad: si hay un equipo en un lugar con un cierto tipo de maquinaria en él, el mismo equipo en otro lugar se comportará de la misma manera. ¿Por qué? Porque una máquina, cuando es analizada por Pedro, tiene exactamente las mismas ecuaciones que la otra, analizada por Juan. Ya que las *ecuaciones* son las mismas, los *fenómenos* son los mismos. Así, la prueba de que un aparato en una nueva posición se comporta igual a como lo hacía en la antigua posición es la misma que la prueba de que las ecuaciones, cuando han sido desplazadas en el espacio, se reproducen. Por lo tanto, diremos que *las leyes de la física son simétricas para desplazamientos de traslación*, simétricas en el sentido de que las leyes no cambian cuando hacemos una translación de nuestras coordenadas. Por supuesto, es intuitivamente bastante obvio que esto es verdad, pero es interesante y entretenido discutir su matemática.

11-3 Rotaciones

La anterior es la primera de una serie de proposiciones cada vez más complicadas relativas a la simetría de una ley física. La siguiente proposición es que no importaría en qué *dirección* escogemos los ejes. En otras palabras, si construimos una parte del equipo en algún lugar y lo observamos funcionar y cerca construimos el mismo tipo de aparato, pero girado en un ángulo respecto al primero, ¿funcionará de la misma manera? ¡Evidentemente que no si es un reloj de péndulo, por ejemplo! Si un reloj de péndulo está vertical, trabaja bien, pero si está inclinado el péndulo cae contra el lado de la caja y nada sucede. El teorema es entonces falso en el caso del reloj de péndulo, a menos que incluyamos la Tierra, que atrae al péndulo. Por lo tanto, podemos hacer una predicción acerca de los relojes de péndulo si creemos en la simetría de las leyes físicas para la rotación: algo más interviene en el funcionamiento de un reloj de péndulo, además de la maquinaria del reloj, algo exterior a él que debemos buscar. También podemos predecir que los relojes de péndulo no trabajarán de la misma manera cuando están ubicados en lugares diferentes en relación con esta misteriosa fuente de asimetría, tal vez la tierra. En verdad, sabemos que un reloj de péndulo en un satélite artificial, por ejemplo, no haría ni tic porque no hay fuerza efectiva, y en Marte iría a una velocidad diferente. Los relojes de péndulo *por cierto* comprenden algo más que la pura maquinaria interna, comprenden algo del exterior. Una vez que reconocemos este factor, vemos que debemos girar la Tierra junto con el aparato. Por supuesto no debemos preocupamos de eso, es fácil de hacer, uno simplemente espera uno o dos minutos y la Tierra gira; entonces el reloj de péndulo hace de nuevo tictac en la nueva posición igual que lo hacía antes. Mientras estemos rotando en el espacio, nuestros ángulos estarán cambiando siempre, en forma absoluta; este cambio parece no molestarnos demasiado, porque en la nueva posición nos parece estar en la misma condición que en la antigua. Esto tiene una cierta tendencia a confundir a uno, porque es verdad que en la nueva posición girada las leyes son iguales que en la posición no girada, pero *no* es verdad *que cuando giramos* una cosa, siga las mismas leyes que cuando no la giramos. Si ejecutamos experimentos suficientemente delicados, podemos decir que la Tierra *está rotando*, pero no que ella *haya rotado.* En otras palabras, no podemos localizar su posición angular, pero podemos decir que está cambiando.

Ahora podemos discutir los efectos de la orientación angular sobre las leyes físicas. Averigüemos si el mismo juego con Juan y Pedro funciona de nuevo. Esta vez, para evitar complicaciones innecesarias, supondremos que Juan y Pedro usan el mismo origen (hemos ya demostrado que los ejes pueden ser movidos por traslación a otro lugar). Supongamos que los ejes de Pedro han rotado con relación a los de Juan en un ángulo θ. Los dos sistemas coordenados se muestran en la figura 11-2, que está restringida a dos dimensiones. Consideremos cualquier punto P que tiene coordenadas (x, y) en el sistema de Juan y (x', y') en el sistema de Pedro. Comenzaremos, como en el caso anterior, por expresar las coordenadas x' y y' en términos de x, y y θ. Para hacerlo así, bajaremos primero perpendiculares desde P a los cuatro ejes y dibujaremos AB perpendicular a PQ. La inspección de la figura muestra que x' puede escribirse como la suma de dos longitudes según el eje x', y y' como la diferencia de dos longitudes según AB. Todas estas longitudes están expresadas en términos de x, y y θ en las ecuaciones (11.5), a las cuales hemos agregado una ecuación para la tercera dimensión.

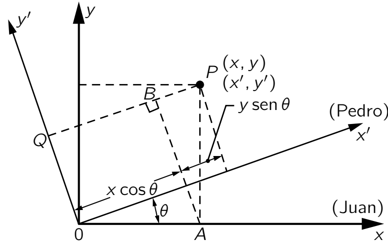


Figura 11-2. Dos sistemas de coordenadas con diferentes orientaciones angulares.

$$\begin{aligned} x' &= x\cos\theta + y\,\mathrm{sen}\,\theta,\\ y' &= y\cos\theta - x\,\mathrm{sen}\,\theta,\\ z' &= z. \end{aligned} \tag{11.5}$$

El próximo paso es analizar la relación de las fuerzas como las ven los dos observadores, siguiendo el mismo método general como antes. Supongamos que una fuerza $\boldsymbol{F}$, la cual ya ha sido analizada como teniendo componentes F_x, y F_y (como la ve Juan), está actuando sobre una partícula de masa m, localizada en el punto P en la figura 11-2. Por simplicidad movamos ambos conjuntos de ejes de modo que el origen esté en P, como se muestra en la figura 11-3. Pedro ve las componentes de $\boldsymbol{F}$ según sus ejes como $F_{x'}$ y $F_{y'}$. F_x tiene componentes según ambos ejes x' y y', análogamente F_y tiene componentes según estos dos ejes. Para expresar $F_{x'}$ en términos de F_x y F_y sumamos estas componentes según el eje x', y en forma parecida podemos expresar $F_{y'}$ en términos de F_x y F_y. Los resultados son

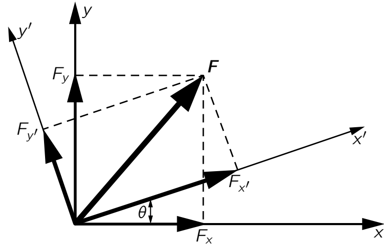


Figura 11-3. Componentes de una fuerza en dos sistemas.

$$\begin{aligned} F_{x'} &= F_x\cos\theta + F_y\,\mathrm{sen}\,\theta,\\ F_{y'} &= F_y\cos\theta - F_x\,\mathrm{sen}\,\theta,\\ F_{z'} &= F_z. \end{aligned} \tag{11.6}$$

Es interesante notar una especie de accidente, que es de extrema importancia: las fórmulas (11.5) y (11.6), para las coordenadas de P y las componentes de $\boldsymbol{F}$, respectivamente, *son de forma idéntica.*

Como antes, se supone que las leyes de Newton son válidas en el sistema de Juan y están expresadas por las ecuaciones (11.1). La pregunta, otra vez, es si Pedro puede aplicar las leyes de Newton —¿serán correctos los resultados para su sistema de ejes rotados?—. En otras palabras, si suponemos que las ecuaciones (11.5) y (11.6) dan una relación de las medidas, ¿es verdad o no que

$$\begin{aligned} m(d^2x'/dt^2) &= F_{x'},\\ m(d^2y'/dt^2) &= F_{y'},\\ m(d^2z'/dt^2) &= F_{z'}? \end{aligned} \tag{11.7}$$

Para comprobar estas ecuaciones, calculamos independientemente los primeros y los segundos miembros y comparamos los resultados. Para calcular los primeros miembros, multiplicamos las ecuaciones (11.5) por m y derivamos dos veces respecto al tiempo, suponiendo que el ángulo θ es constante. Esto da

$$\begin{aligned} m(d^2x'/dt^2) &= m(d^2x/dt^2)\cos\theta + m(d^2y/dt^2)\,\mathrm{sen}\,\theta,\\ m(d^2y'/dt^2) &= m(d^2y/dt^2)\cos\theta - m(d^2x/dt^2)\,\mathrm{sen}\,\theta,\\ m(d^2z'/dt^2) &= m(d^2z/dt^2). \end{aligned} \tag{11.8}$$

Calculamos los segundos miembros de las ecuaciones (11.7) por sustitución de las ecuaciones (11.1) en las ecuaciones (11.6). Esto da

$$\begin{aligned} F_{x'} &= m(d^2x/dt^2)\cos\theta + m(d^2y/dt^2)\operatorname{sen}\theta,\\ F_{y'} &= m(d^2y/dt^2)\cos\theta - m(d^2x/dt^2)\operatorname{sen}\theta,\\ F_{z'} &= m(d^2z/dt^2). \end{aligned} \tag{11.9}$$

¡Miren! Los segundos miembros de las ecuaciones (11.8) y (11.9) son idénticos; así que concluimos que si las leyes de Newton son correctas en un conjunto de ejes, serán también válidas para cualquier otro conjunto de ejes. Este resultado, que ha sido ahora establecido tanto para traslación como para rotación de ejes, tiene ciertas consecuencias: primero, nadie puede pretender que sus ejes particulares sean únicos; pero, por supuesto, pueden ser más *convenientes* para algunos problemas particulares. Por ejemplo, es útil tener la gravedad a lo largo de un eje, pero esto no es físicamente necesario. Segundo, significa que cualquier equipo que sea completamente autosuficiente, con todo el equipo generador de fuerza completamente dentro del aparato, trabajaría igual cuando se girara en un ángulo.

11-4 Vectores

No solamente las leyes de Newton, sino también las otras leyes de la física, hasta donde sabemos hoy día, tienen las dos propiedades que llamamos invariancia (o simetría) frente a traslación y rotación de ejes. Estas propiedades son tan importantes que una técnica matemática ha sido desarrollada para aprovecharlas para escribir y usar leyes físicas.

El análisis precedente implicó considerable trabajo matemático tedioso. Para reducir los detalles a un mínimo en el análisis de tales problemas, se ha diseñado una maquinaria matemática muy poderosa. Este sistema, llamado *análisis vectorial*, da el título a este capítulo; sin embargo, estrictamente hablando, éste es un capítulo sobre la simetría de las leyes físicas. Por los métodos del análisis anterior pudimos hacer todo lo necesario para obtener los resultados que buscábamos, pero en la práctica nos gustaría hacer las cosas más fácil y rápidamente; por lo tanto, empleamos la técnica vectorial.

Comenzamos notando algunas características de dos tipos de cantidades que son importantes en la física. (Realmente hay más de dos, pero empecemos con dos.) Uno de ellos, como el número de papas en un saco, lo llamamos una cantidad ordinaria, o una cantidad no dirigida o un *escalar*. La temperatura es un ejemplo de tal cantidad. Otras cantidades que son importantes en física tienen dirección, por ejemplo la velocidad: debemos estar al tanto de hacia dónde un cuerpo se dirige, no sólo de su rapidez. El momentum y la fuerza también tienen dirección, como lo tiene el desplazamiento: cuando alguien camina de un lugar a otro en el espacio, podemos estar al tanto de cuánto se alejó, pero si queremos también saber hacia *dónde* fue, tenemos que especificar una dirección.

Todas las cantidades que tienen una dirección, como un paso en el espacio, se llaman *vectores*.

Un vector es tres números. A fin de representar un paso en el espacio, digamos desde el origen a algún punto particular P cuya ubicación es (x, y, z), necesitamos realmente tres números, pero vamos a inventar un único símbolo matemático, $\boldsymbol{r}$, que es distinto a todos los otros símbolos matemáticos, que hemos usado hasta ahora.* *No* es un número único, representa a *tres* números: x, y y z. Significa tres números, pero realmente sólo *esos* tres números, porque si usáramos un sistema de coordenadas diferentes, los tres números cambiarían a x', y' y z'. Sin embargo, queremos mantener nuestra matemática simple y así vamos a usar el *mismo símbolo* para representar los tres números (x, y, z) y los tres números (x', y', z'). Esto es, usamos el mismo símbolo para representar al primer conjunto de tres números para un sistema de coordenadas, y al segundo conjunto de tres

* En letra de imprenta, los vectores se representan por tipos en negrita; en manuscrito usa una flecha: $\vec{r}$.

números, si estamos usando el otro sistema de coordenadas. Esto tiene la ventaja que cuando cambiamos el sistema de coordenadas no tenemos que cambiar las letras de nuestras ecuaciones. Si escribimos una ecuación en términos de x, y, z, y luego usamos otro sistema, tenemos que cambiar a x', y', z' pero escribiremos sólo $\boldsymbol{r}$, con la convención de que representa (x, y, z) si usamos un conjunto de ejes, o (x', y', z') si usamos otro conjunto de ejes, y así sucesivamente. Los tres números que describen la cantidad en un sistema de coordenadas dado se llaman las *componentes* del vector en la dirección de los ejes coordenados de ese sistema. Esto es, usamos el mismo símbolo para las tres letras que corresponden al *mismo objeto, como se ven desde diferentes ejes*. El hecho mismo de que podamos decir "el mismo objeto" implica una intuición física acerca de la realidad de un paso en el espacio, que es independiente de las componentes en término de las cuales lo medimos. Así el símbolo $\boldsymbol{r}$ representará la misma cosa, sin importar como giremos los ejes.

Ahora supongamos que hay otra cantidad física dirigida, cualquier otra cantidad, que tiene también tres números asociados a ella, como la fuerza, y estos tres números se cambian a otros tres números por alguna regla matemática, si cambiamos los ejes. Debe ser la misma regla que cambia (x, y, z) a (x', y', z'). En otras palabras, cualquier cantidad física asociada con tres números que se transforman como lo hacen las componentes de un paso en el espacio es un vector. Una ecuación como

$$\boldsymbol{F} = \boldsymbol{r}$$

sería así válida en *cualquier* sistema de coordenadas si lo fuera en uno. Esta ecuación, por supuesto, representa las tres ecuaciones:

$$F_x = x, \qquad F_y = y, \qquad F_z = z,$$

o, equivalentemente,

$$F_{x'} = x', \qquad F_{y'} = y', \qquad F_{z'} = z'.$$

El hecho de que una relación física pueda ser expresada como una ecuación vectorial nos asegura que la relación no cambia por una simple rotación del sistema de coordenadas. Ésa es la razón de que los vectores sean tan útiles en la física. Ahora examinaremos algunas propiedades de los vectores. Como ejemplo de vectores podemos mencionar la velocidad, el momentum, la fuerza y la aceleración. Para muchos propósitos es conveniente representar una cantidad vectorial por una flecha que indica la dirección en la cual está actuando. ¿Por qué podemos representar la fuerza, digamos, por una flecha? Porque tiene las mismas propiedades de transformación matemática que "un paso en el espacio". Así, lo representamos en un diagrama como si fuera un paso, usando una escala tal que una unidad de fuerza, o un newton, corresponda a cierta longitud conveniente. Una vez que hayamos hecho esto, todas las fuerzas pueden estar representadas como longitudes, porque una ecuación como

$$\boldsymbol{F} = k\boldsymbol{r},$$

donde k es alguna constante, es una ecuación perfectamente legítima. Así siempre podemos representar fuerzas por segmentos, lo cual es muy conveniente, porque una vez que hemos dibujado el segmento no necesitamos más los ejes. Por consiguiente, podemos calcular rápidamente las tres componentes a medida que cambian cuando giran los ejes porque esto es sólo un problema geométrico.

11-5 Álgebra vectorial

Debemos ahora describir las leyes, o reglas, para combinar vectores de diferentes maneras. La primera de estas combinaciones es la *adición* de dos vectores: supongamos que $\boldsymbol{a}$ es un vector el cual en algún sistema de coordenadas particular tiene las tres componentes (a_x, a_y, a_z), y que $\boldsymbol{b}$ es otro vector que tiene las tres componentes (b_x, b_y, b_z). Inventemos ahora tres nuevos números $(a_x + b_x, a_y + b_y, a_z + b_z)$. ¿Forman éstos un vector? "Bien", podemos decir, "son

tres números, y cada tres números son un vector". ¡No, *no* cada tres números forman un vector! Para que esto sea un vector, no sólo debe haber tres números, sino que éstos deben estar asociados a un sistema de coordenadas de tal manera que si giramos el sistema de coordenadas, los tres números "se revuelven", "mezclándose" entre sí, según las leyes precisas que ya hemos descrito. Así que la pregunta es, si ahora rotamos el sistema de coordenadas tal que (a_x, a_y, a_z) llegue a ser $(a_{x'}, a_{y'}, a_{z'})$ y (b_x, b_y, b_z) llegue a ser $(b_{x'}, b_{y'}, b_{z'})$, ¿qué llegará a ser $(a_x + b_x, a_y + b_y, a_z, +b_z)$? ¿Llegan a ser $(a_{x'} + b_{x'}, a_{y'} + b_{y'}, a_{z'} + b_{z'})$ o no? La respuesta es sí, por supuesto, porque el prototipo de transformaciones de la ecuación (11.5) constituye lo que llamamos una transformación *lineal*. Si aplicamos esas transformaciones a a_x y b_x para obtener $a_{x'} + b_{x'}$, encontramos que la transformada $a_x + b_x$ es por cierto la misma que $a_{x'} + b_{x'}$. Cuando $\boldsymbol{a}$ y $\boldsymbol{b}$ se "suman" en este sentido, formarán un vector que podemos llamar $\boldsymbol{c}$. Escribiremos esto como

$$\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}.$$

Ahora $\boldsymbol{c}$ tiene la interesante propiedad

$$\boldsymbol{c} = \boldsymbol{b} + \boldsymbol{a},$$

como podemos ver inmediatamente a partir de sus componentes. Así también,

$$\boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) = (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c}.$$

Podemos sumar los vectores en cualquier orden.

¿Cuál es el significado geométrico de $\boldsymbol{a} + \boldsymbol{b}$? Supongamos que $\boldsymbol{a}$ y $\boldsymbol{b}$ fueran representados por segmentos sobre un pedazo de papel. ¿Qué sería $\boldsymbol{c}$? Esto se muestra en la figura 11-4. Observamos que podemos sumar las componentes de $\boldsymbol{b}$ a aquellas de $\boldsymbol{a}$ más convenientemente si colocamos el rectángulo que representa las componentes de $\boldsymbol{b}$ junto al que representa las componentes de $\boldsymbol{a}$ de la manera indicada. Ya que $\boldsymbol{b}$ se "ajusta" precisamente en su rectángulo, como lo hace $\boldsymbol{a}$ en su rectángulo, esto es lo mismo que colocar la "cola" de $\boldsymbol{b}$ sobre la "cabeza" de $\boldsymbol{a}$ siendo el vector $\boldsymbol{c}$ desde la "cola" de $\boldsymbol{a}$ a la "cabeza" de $\boldsymbol{b}$. Por supuesto, si sumamos $\boldsymbol{a}$ a $\boldsymbol{b}$ de la otra manera, pondríamos la "cola" de $\boldsymbol{a}$ sobre la "cabeza" de $\boldsymbol{b}$, y por las propiedades geométricas de los paralelogramos obtendríamos el mismo resultado para $\boldsymbol{c}$. Nótese que los vectores se pueden sumar en esta forma sin hacer referencia a ejes de coordenadas.

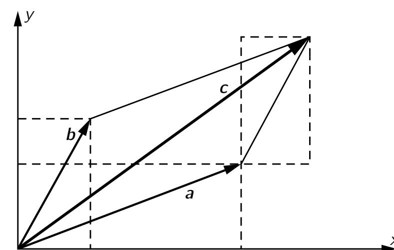


Figura 11-4. La suma de vectores.

Supongan que multiplicamos un vector por un número α, ¿qué significa esto? Lo *definimos* para indicar un nuevo vector cuyas componentes son αa_x, αa_y y αa_z. Lo dejamos como problema para que el estudiante demuestre que *es* un vector.

Consideremos ahora la sustracción de vectores. Podemos definir a la sustracción de la misma manera como la adición, pero en lugar de sumar restamos las componentes. O podríamos definir la sustracción definiendo un vector negativo, $-\boldsymbol{b} = -1\boldsymbol{b}$, y entonces sumaríamos las componentes. Se llega al mismo resultado. Este resultado se muestra en la figura 11-5. Esta figura muestra $\boldsymbol{d} = \boldsymbol{a} - \boldsymbol{b} = \boldsymbol{a} + (-\boldsymbol{b})$; también notamos que la diferencia $\boldsymbol{a} - \boldsymbol{b}$ se puede encontrar muy fácilmente a partir de $\boldsymbol{a}$ y $\boldsymbol{b}$ usando la relación equivalente $\boldsymbol{a} = \boldsymbol{b} + \boldsymbol{d}$. Así, la diferencia es aún más fácil de encontrar que la suma: ¡sólo dibujamos el vector desde $\boldsymbol{b}$ hasta $\boldsymbol{a}$ para obtener $\boldsymbol{a} - \boldsymbol{b}$!

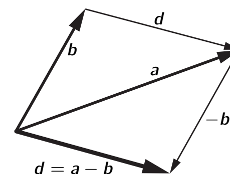


Figura 11-5. La resta de vectores.

A continuación discutimos la velocidad. ¿Por qué la velocidad es un vector? Si la posición está dada por las tres coordenadas (x, y, z), ¿qué es la velocidad? La velocidad está dada por dx/dt, dy/dt y dz/dt. ¿Es esto un vector, o no? Podemos averiguarlo derivando las expresiones en la ecuación (11.5) para encontrar si la dx'/dt se *transforma* de la manera correcta. Vemos que las componentes dx/dt y dy/dt *se* transforman de acuerdo a las mismas leyes que x y y, y por lo tanto la derivada temporal es un vector. Así, pues, la velocidad *es* un vector. Podemos escribir la velocidad de una manera interesante como

$$\boldsymbol{v} = d\boldsymbol{r}/dt.$$

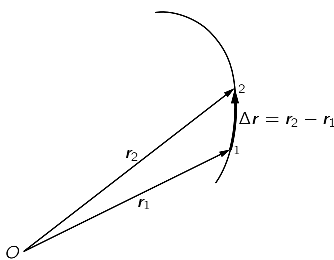


Figura 11-6. El desplazamiento de una partícula en un pequeño intervalo de tiempo $\Delta t = t_2 - t_1$.

Lo que la velocidad es y por qué es un vector, también se puede entender en forma más gráfica: ¿hasta dónde se moverá una partícula en un tiempo corto Δt? Respuesta: $\Delta \boldsymbol{r}$; así, si una partícula está "aquí" en un instante y "allá" en otro instante, entonces la diferencia vectorial de las posiciones $\Delta \boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1$, que está en la dirección del movimiento mostrado en la figura 11-6, dividido entre el intervalo de tiempo $\Delta t = t_2 - t_1$, es el vector "velocidad media".

En otras palabras, por vector velocidad queremos indicar el límite para Δt tendiendo a 0, de la diferencia entre los radios vectores en el tiempo $t + \Delta t$ y el tiempo t, dividido entre Δt:

$$\boldsymbol{v} = \lim_{\Delta t \to 0} (\Delta \boldsymbol{r}/\Delta t) = d\boldsymbol{r}/dt. \tag{11.10}$$

Así, la velocidad es un vector, porque es la diferencia de dos vectores. Es también la definición correcta de velocidad, porque sus componentes son dx/dt, dy/dt y dz/dt. En efecto, vemos conforme a estos razonamientos que si derivamos *cualquier* vector respecto al tiempo, formamos un nuevo vector. Así, tenemos varias maneras de formar nuevos vectores: *(*1*)* multiplicar por una constante, *(*2*)* derivar con respecto al tiempo y *(*3*)* sumar o restar dos vectores.

11-6 Leyes de Newton en notación vectorial

Para escribir las leyes de Newton en forma vectorial, tenemos que avanzar un paso más aún y definir el vector aceleración. Éste es la derivada temporal del vector velocidad y es fácil de demostrar que sus componentes son las segundas derivadas de x, y y z respecto a t:

$$\boldsymbol{a} = \frac{d\boldsymbol{v}}{dt} = \left(\frac{d}{dt}\right)\left(\frac{d\boldsymbol{r}}{dt}\right) = \frac{d^2\boldsymbol{r}}{dt^2}, \tag{11.11}$$

$$a_x = \frac{dv_x}{dt} = \frac{d^2x}{dt^2}, \qquad a_y = \frac{dv_y}{dt} = \frac{d^2y}{dt^2}, \qquad a_z = \frac{dv_z}{dt} = \frac{d^2z}{dt^2}. \tag{11.12}$$

Con esta definición, entonces, las leyes de Newton pueden escribirse de esta manera:

$$m\boldsymbol{a} = \boldsymbol{F} \tag{11.13}$$

o

$$m(d^2\boldsymbol{r}/dt^2) = \boldsymbol{F}. \tag{11.14}$$

Ahora el problema de demostrar la invariancia de las leyes de Newton frente a una rotación de coordenadas es éste: demostrar que $\boldsymbol{a}$ es un vector; esto ya lo hemos hecho. Demostrar que $\boldsymbol{F}$ es un vector; *suponemos* que lo es. Así, si la fuerza es un vector, entonces, ya que sabemos que la aceleración es un vector, la ecuación (11.13) será la misma en cualesquier sistemas de coordenadas. Escribirla en una forma que no contenga explícitamente las x, y y z tiene la ventaja de que a partir de ahora no necesitamos escribir *tres* leyes cada vez que escribimos las ecuaciones de Newton u otras leyes de la física. Escribimos lo que parece ser *una* ley; pero realmente, por supuesto, representa las tres leyes para cualquier conjunto particular de ejes, porque cualquier ecuación vectorial encierra la afirmación de que *cada una de las componentes es igual.*

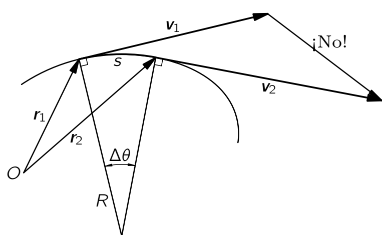


Figura 11-7. Una trayectoria curva.

El hecho de que la aceleración es la derivada temporal del vector velocidad nos ayuda a calcular la aceleración en algunas circunstancias bastante complicadas. Supongamos, por ejemplo, que una partícula se está moviendo sobre alguna curva complicada (figura 11-7) y que, en un instante t_1 dado, tenga una cierta velocidad $\boldsymbol{v}_1$, pero que cuando vamos a otro instante t_2 un poco después, tiene una velocidad diferente $\boldsymbol{v}_2$. ¿Cuál es la aceleración? Respuesta: la aceleración es la diferencia en la velocidad dividida por el pequeño intervalo de tiempo, así que necesitamos la diferencia de las dos velocidades. ¿Cómo obtenemos la diferencia entre las velocidades? Para restar dos vectores, trazamos el vector por los extremos de $\boldsymbol{v}_2$ y $\boldsymbol{v}_1$; esto es, dibujamos $\Delta \boldsymbol{v}$ como la diferencia de los

dos vectores. ¿Correcto? ¡*No*! ¡Eso solamente sucede cuando las *colas* de los vectores están en el mismo lugar! ¡No tiene sentido si movemos el vector a algún otro lugar y enseguida dibujamos una línea para unir los extremos, así que ¡cuidado! Tenemos que dibujar un nuevo diagrama para restar los vectores. En la figura 11-8, $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ están ambas dibujadas paralelas e iguales a sus contrapartes en la figura 11-7, y ahora podemos discutir la aceleración. Por supuesto, la aceleración es simplemente $\Delta\boldsymbol{v}/\Delta t$. Es interesante notar que podemos componer la diferencia de velocidad de dos partes; podemos pensar que la aceleración tiene *dos componentes*, $\Delta v_{\parallel}$ en la dirección tangente a la trayectoria y $\Delta v_{\perp}$ perpendicular a la trayectoria, como se indica en la figura 11-8. La aceleración tangente a la trayectoria es, por supuesto, justamente el cambio de *longitud* del vector, es decir, el cambio en la *velocidad* v:

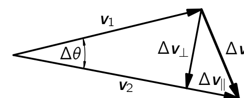


Figura 11-8. Diagrama para calcular la aceleración.

$$a_{\parallel} = dv/dt. \tag{11.15}$$

La otra componente de la aceleración, con ángulos correctos a la curva, es fácil de calcular, usando las figuras 11-7 y 11-8. En el tiempo corto Δt sea $\Delta\theta$ el pequeño ángulo que representa el cambio en ángulo entre $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$. Si el módulo de la velocidad se llama v, entonces, por supuesto,

$$\Delta v_{\perp} = v\,\Delta\theta$$

y la aceleración a será:

$$a_{\perp} = v\,(\Delta\theta/\Delta t).$$

Ahora necesitamos conocer $\Delta\theta/\Delta t$, que se puede encontrar de esta manera: si en un momento dado la curva se aproxima a una circunferencia de cierto radio R, entonces en un tiempo Δt la distancia s es, por supuesto, $v\Delta t$, donde v es el módulo de la velocidad.

$$\Delta\theta = (v\,\Delta t)/R, \qquad \text{o} \qquad \Delta\theta/\Delta t = v/R.$$

Por lo tanto, encontramos

$$a_{\perp} = v^2/R, \tag{11.16}$$

como hemos visto antes.

11-7 Producto escalar de vectores

Examinemos ahora un poco más las propiedades de los vectores. Es fácil ver que la *longitud* de un paso en el espacio sería la misma en cualquier sistema de coordenadas. Esto es, si un paso particular r se representa por x, y, z, en un sistema de coordenadas, y por x', y', z', en otro sistema de coordenadas, seguramente la distancia $r = |\boldsymbol{r}|$ en ambos serían la misma. Ahora

$$r = \sqrt{x^2 + y^2 + z^2}$$

y también

$$r' = \sqrt{x'^2 + y'^2 + z'^2}.$$

Por lo tanto, lo que queremos verificar es que estas dos cantidades son iguales. Es mucho más conveniente no preocuparse de la raíz cuadrada; así hablemos del cuadrado de la distancia; esto es, averigüemos si

$$x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2. \tag{11.17}$$

Sería bueno que fuera así —y si sustituimos la ecuación (11.5) encontramos que realmente lo es—. Así, pues, vemos que hay otras clases de ecuaciones que son válidas para dos sistemas de coordenadas cualesquiera.

Alguno nuevo está involucrado. Podemos producir una nueva cantidad, una función de x, y y z, llamada *función escalar*, una cantidad que no tiene dirección, sino que es la misma en ambos sistemas. A partir de un vector podemos construir

un escalar. Tenemos que encontrar una regla general para eso. Es claro lo que la regla es para el caso recién considerado: sumar los cuadrados de las componentes. Ahora definamos algo nuevo, que llamaremos $\boldsymbol{a}\cdot\boldsymbol{a}$. Esto no es un vector, sino un escalar; es un número que es el mismo en todos los sistemas de coordenadas y se define como la suma de los cuadrados de las tres componentes del vector:

$$\boldsymbol{a}\cdot\boldsymbol{a} = a_x^2 + a_y^2 + a_z^2. \tag{11.18}$$

Ahora ustedes dirán: "¿Pero con qué ejes?" No depende de los ejes; la respuesta es la misma en *todo* conjunto de ejes. Así tenemos un nuevo *tipo* de cantidad, un nuevo *invariante* o *escalar*, producido por un vector "al cuadrado". Si ahora definimos la siguiente cantidad para dos vectores cualesquiera $\boldsymbol{a}$ y $\boldsymbol{b}$:

$$\boldsymbol{a}\cdot\boldsymbol{b} = a_x b_x + a_y b_y + a_z b_z, \tag{11.19}$$

encontramos que esta cantidad, calculada en sistemas con prima y sin prima, también sigue siendo la misma. Para probarlo, notemos que es válido para $\boldsymbol{a}\cdot\boldsymbol{a}$, $\boldsymbol{b}\cdot\boldsymbol{b}$ y $\boldsymbol{c}\cdot\boldsymbol{c}$, donde $\boldsymbol{c}=\boldsymbol{a}+\boldsymbol{b}$. Por lo tanto, la suma de los cuadrados $(a_x+b_x)^2+(a_y+b_y)^2+(a_z+b_z)^2$ será invariante:

$$(a_x+b_x)^2+(a_y+b_y)^2+(a_z+b_z)^2 = (a_{x'}+b_{x'})^2 + (a_{y'}+b_{y'})^2+(a_{z'}+b_{z'})^2. \tag{11.20}$$

Si ambos miembros de esta ecuación se desarrollan, habría productos mixtos del mismo tipo que aparece en la ecuación (11.19), así como las sumas de los cuadrados de las componentes de $\boldsymbol{a}$ y $\boldsymbol{b}$. La invariancia de términos de la forma de la ecuación (11.18) deja entonces también invariantes los términos con productos mixtos (11.19).

La cantidad $\boldsymbol{a}\cdot\boldsymbol{b}$ se llama *producto escalar* de dos vectores $\boldsymbol{a}$ y $\boldsymbol{b}$, y tiene propiedades muy útiles e interesantes. Por ejemplo, se prueba fácilmente que

$$\boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c}) = \boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}. \tag{11.21}$$

Además, existe una manera geométrica simple para calcular $\boldsymbol{a}\cdot\boldsymbol{b}$, sin tener que calcular las componentes de $\boldsymbol{a}$ y $\boldsymbol{b}$: $\boldsymbol{a}\cdot\boldsymbol{b}$ es el producto del largo de $\boldsymbol{a}$ por el largo de $\boldsymbol{b}$ por el coseno del ángulo entre ellos. ¿Por qué? Supongamos que escogemos un sistema especial de coordenadas en el cual el eje x queda a lo largo de $\boldsymbol{a}$; en ese caso la única componente no cero de $\boldsymbol{a}$ es a_x, que es, por supuesto, el largo total de $\boldsymbol{a}$. Así la ecuación (11.19) se reduce a $\boldsymbol{a}\cdot\boldsymbol{b}=a_x b_x$ para este caso, y éste es el largo de $\boldsymbol{a}$ por la componente de $\boldsymbol{b}$ en la dirección de $\boldsymbol{a}$, esto es, $b\cos\theta$:

$$\boldsymbol{a}\cdot\boldsymbol{b} = ab\cos\theta.$$

Por lo tanto, en este sistema especial de coordenadas, hemos probado que $\boldsymbol{a}\cdot\boldsymbol{b}$ es el largo de $\boldsymbol{a}$ por el largo de $\boldsymbol{b}$ por el $\cos\theta$. *Pero si esto es verdad en un sistema de coordenadas, lo es en todos*, porque $\boldsymbol{a}\cdot\boldsymbol{b}$ es independiente del sistema de coordenadas; ése es nuestro razonamiento.

¿Cuál es la ventaja del producto escalar? ¿Existen casos en la física donde lo necesitamos? Sí, lo necesitamos todo el tiempo. Por ejemplo, en el capítulo 4 la energía cinética se indicó por $\frac{1}{2}mv^2$, pero si el objeto se está moviendo en el espacio sería el cuadrado de la velocidad en la dirección x, en la dirección y y en la dirección z, y así la fórmula para la energía cinética de acuerdo con el cálculo vectorial es

$$\text{EC} = \tfrac{1}{2}m(\boldsymbol{v}\cdot\boldsymbol{v}) = \tfrac{1}{2}m(v_x^2+v_y^2+v_z^2). \tag{11.22}$$

La energía no tiene dirección. El momentum tiene dirección, es un vector y es la masa por el vector velocidad.

Otro ejemplo de un producto escalar es el trabajo realizado por una fuerza cuando algo se empuja desde un lugar a otro. No hemos definido aún el trabajo, pero éste es equivalente al cambio de energía como al levantar pesos cuando una fuerza $\boldsymbol{F}$ actúa una distancia $\boldsymbol{s}$:

$$\text{Trabajo} = \boldsymbol{F} \cdot \boldsymbol{s}. \tag{11.23}$$

Algunas veces es muy conveniente hablar de las componentes de un vector en cierta dirección (digamos la dirección vertical, porque ésa es la dirección de la gravedad). Para tales propósitos es útil inventar lo que llamamos un *vector unitario* en la dirección que queremos estudiar. Por un vector unitario entendemos un vector cuyo producto escalar consigo mismo es igual a la unidad. Llamemos a este vector $\boldsymbol{i}$; entonces $\boldsymbol{i} \cdot \boldsymbol{i} = 1$. Entonces si queremos la componente de algún vector en la dirección de $\boldsymbol{i}$, vemos que el producto escalar $\boldsymbol{a} \cdot \boldsymbol{i}$ será $a \cos 0$, es decir, la componente de $\boldsymbol{a}$ en la dirección de $\boldsymbol{i}$. Ésta es una bonita manera de obtener la componente. En realidad, nos permite obtener *todas* las componentes y escribir una fórmula sorprendente. Supongan que en un sistema dado de coordenadas, x, y y z, inventamos tres vectores: $\boldsymbol{i}$, un vector unitario en la dirección x; $\boldsymbol{j}$, un vector unitario en la dirección y y $\boldsymbol{k}$, un vector unitario en la dirección z. Notemos primero que $\boldsymbol{i} \cdot \boldsymbol{i} = 1$. ¿Qué es $\boldsymbol{i} \cdot \boldsymbol{j}$? Cuando dos vectores son perpendiculares, su producto escalar es cero. Así

$$\begin{aligned}
&\boldsymbol{i} \cdot \boldsymbol{i} = 1 \\
&\boldsymbol{i} \cdot \boldsymbol{j} = 0 \qquad \boldsymbol{j} \cdot \boldsymbol{j} = 1 \\
&\boldsymbol{i} \cdot \boldsymbol{k} = 0 \qquad \boldsymbol{j} \cdot \boldsymbol{k} = 0 \qquad \boldsymbol{k} \cdot \boldsymbol{k} = 1
\end{aligned} \tag{11.24}$$

Ahora, con estas definiciones, un vector cualquiera puede escribirse de esta manera:

$$\boldsymbol{a} = a_x \boldsymbol{i} + a_y \boldsymbol{j} + a_z \boldsymbol{k}. \tag{11.25}$$

De este modo podemos ir desde las componentes de un vector al vector mismo.

Esta discusión de los vectores de ningún modo es completa. Sin embargo, en vez de tratar de entrar ahora más profundamente en el tema, aprenderemos primero a usar en situaciones físicas algunas de las ideas discutidas hasta aquí. Entonces, cuando hayamos dominado debidamente este material básico, encontraremos más fácil penetrar más profundamente en el tema sin llegar a confundirnos demasiado. Encontraremos más tarde que es útil definir otra clase de producto de dos vectores, llamado producto vectorial, y que se escribe $\boldsymbol{a} \times \boldsymbol{b}$. Sin embargo, emprenderemos una discusión de tales cosas en un capítulo posterior.

12

Características de la fuerza

12-1 ¿Qué es una fuerza?

A pesar de que es interesante y vale la pena estudiar las leyes físicas, sencillamente porque nos ayudan a comprender y utilizar la naturaleza, uno debería detenerse de vez en cuando y pensar: "¿Qué significan realmente?" El significado de cualquier afirmación es un tema que ha interesado e inquietado a los filósofos desde tiempos inmemoriales, y el significado de las leyes físicas es aún más interesante, porque generalmente se cree que estas leyes representan alguna forma de conocimiento real. El significado del conocimiento es un problema profundo en la filosofía y es siempre importante preguntar: "¿Qué significa?"

Preguntemos: "¿Cuál es el significado de las leyes físicas de Newton, que escribimos como $F = ma$? ¿Cuál es el significado de fuerza, masa y aceleración?" Bueno, intuitivamente podemos percibir el significado de masa, y podemos *definir* aceleración si conocemos el significado de posición y tiempo. No discutiremos los significados de estos términos, sino que nos concentraremos en el nuevo concepto de *fuerza*. La respuesta es igualmente simple: "Si un cuerpo está acelerando, entonces una fuerza actúa sobre él". Eso es lo que dicen las leyes de Newton, de manera que la más precisa y bella definición de fuerza imaginable puede decir sencillamente que fuerza es la masa de un objeto multiplicada por la aceleración. Supongan que tenemos una ley que diga que la conservación del momentum es válida si la suma de todas las fuerzas externas es cero, entonces surge la pregunta: "¿Qué *significa* que la suma de todas las fuerzas externas sea cero?" Una forma cómoda de definir esa afirmación sería: "Cuando el momentum total es una constante, la suma de las fuerzas externas es cero". Debe haber algo erróneo en eso, porque, simplemente, no dice nada nuevo. Si hemos descubierto una ley fundamental que afirma que la fuerza es igual al producto de la masa por la aceleración, y después se *define* que la fuerza es masa por aceleración, no hemos averiguado nada. Podríamos también definir la fuerza diciendo que un objeto en movimiento sobre el cual no actúan fuerzas continúa moviéndose con velocidad constante en línea recta. Entonces, si observamos que un objeto *no* se mueve en línea recta con velocidad constante, podríamos decir que actúa una fuerza sobre él. Ahora, tales cosas verdaderamente no pueden ser el contenido de la física, porque son definiciones que van en círculo vicioso. Sin embargo, la aseveración newtoniana precedente parece ser una definición muy precisa de fuerza, y una que atrae al matemático. Sin embargo, es totalmente inútil, porque de una definición no puede hacerse predicción alguna. Uno puede sentarse todo un día en un sillón y definir tales palabras a voluntad, pero averiguar qué sucede cuando dos esferas se empujan mutuamente o cuando se cuelga un peso de un resorte es completamente distinto, porque la forma en que se *comportan* los cuerpos es algo que queda completamente fuera de toda definición.

Por ejemplo, si quisiéramos decir que un objeto que se deja en un lugar mantiene su posición y no se mueve, entonces cuando vemos que algo se desplaza podríamos decir que se debe a una "güerza" —una "güerza" es la velocidad de cambio de posición—. Ahora tenemos una maravillosa ley nueva, todo permanece en reposo, a menos que actúe una "güerza". Se dan cuenta de que esto sería análogo a la anterior definición de fuerza y no contendría información alguna. El contenido real de las leyes de Newton es éste, que se supone que la fuerza tiene algunas *propiedades independientes*, además de la ley $F = ma$; pero las propiedades *específicas* independientes que tiene la fuerza no fueron descritas totalmente por Newton o por persona alguna y, por consiguiente, la

ley física $F = ma$ es una ley incompleta. Implica que si estudiamos la masa por la aceleración, y denominamos fuerza al producto, esto es, si estudiamos las características de la fuerza como un programa de interés, entonces encontraremos que las fuerzas tienen algo de simplicidad, la ley es un buen programa para analizar la naturaleza, es una sugerencia de que las fuerzas serán sencillas.

Ahora bien, el primer ejemplo de tales fuerzas fue la ley completa de gravitación, que fue dada por Newton y al enunciar la ley contestó la pregunta: "¿Qué es la fuerza?" Si no hubiera nada más que gravitación, entonces la combinación de esta ley y la ley de la fuerza (segunda ley del movimiento) sería una teoría completa, pero existe mucho más que gravitación y queremos emplear las leyes de Newton en muchas situaciones diferentes. Por consiguiente, para poder continuar, tenemos que decir algo acerca de las propiedades de la fuerza.

Por ejemplo, al tener que ver con las fuerzas, siempre se supone tácitamente que la fuerza es igual a cero, a menos que se encuentre presente un cuerpo físico, esto es, si encontramos una fuerza distinta de cero, también encontramos que existe algo a su alrededor que es una fuente de tal fuerza. Esta suposición es enteramente diferente del caso de la "güerza", que introdujimos anteriormente. Una de las características más importantes de una fuerza es que tiene un origen material, y esto *no* es meramente una definición.

Newton también dio una regla acerca de la fuerza, las fuerzas que interactúan entre cuerpos son iguales y opuestas —acción igual a reacción—, esta regla resulta no ser exactamente válida. En realidad, la ley $F = ma$ no es exactamente válida, si fuera una definición deberíamos decir que *siempre* es verdadera, pero no lo es.

El estudiante puede objetar: "No me gusta esta imprecisión, me gustaría tener todo definido en forma exacta; en realidad, algunos libros dicen que toda ciencia es una materia exacta, en la que *todo* está definido". Si insisten en una definición precisa de fuerza, ¡nunca la tendrán! Primero, porque la segunda ley de Newton no es exacta, y segundo, porque para comprender las leyes físicas deben comprender que todas son alguna forma de aproximación.

Toda idea sencilla es aproximada; como ilustración, consideren un objeto... ¿qué es un objeto? Los filósofos siempre dicen: "Bien, tome simplemente una silla, por ejemplo". En el momento que dicen eso, ustedes se dan cuenta que no saben más de lo que están hablando. ¿Qué *es* una silla? Bueno, una silla es algo que está por ahí... ¿cierto?, ¿cuán cierto? Los átomos se están evaporando de ella cada cierto tiempo —no muchos átomos, pero algunos—, mugre cae sobre ella y se disuelve en la pintura, de manera que definir una silla con precisión, decir exactamente qué átomos son silla y qué átomos son aire, o qué átomos son mugre, o qué átomos son pintura que pertenecen a la silla, es imposible. De manera que la masa de una silla puede definirse sólo en forma aproximada. Del mismo modo, definir la masa de un objeto aislado es imposible, porque no hay objetos simples y aislados en el mundo: todo objeto es una mezcla de muchas cosas, de manera que podemos trabajar con ellos sólo como una serie de aproximaciones e idealizaciones.

La clave está en las idealizaciones. Es una excelente aproximación, de tal vez de una parte en 10^{10}, el número de átomos de la silla no cambia en un minuto y, si no somos demasiado precisos, podemos idealizar la silla como una cosa definida. En igual forma aprenderemos acerca de las características de una fuerza de una manera ideal, si no somos demasiado precisos. Se puede estar insatisfecho con la visión aproximada de la naturaleza que la física procura obtener (el intento es siempre aumentar la exactitud de la aproximación), y se puede preferir una definición matemática, pero las definiciones matemáticas no pueden funcionar nunca en el mundo real. Una definición matemática puede ser buena para los matemáticos, en que toda la lógica puede seguirse completamente, pero el mundo físico es complejo, como lo hemos indicado en varios ejemplos, como los de las olas del mar y una copa de vino. Cuando tratamos de aislar partes, hablar de una masa, el vino y la copa, ¿cómo podemos saber cuál es cuál, si uno se disuelve en el otro? Las fuerzas que actúan sobre una sola cosa ya encierran aproximaciones, y si tenemos un sistema de razonamiento acerca del mundo real, entonces ese sistema, al menos hoy en día, debe contener aproximaciones de alguna especie.

Este sistema es bastante diferente del caso de la matemática, en que todo puede ser definido y entonces no *sabemos* de qué estamos hablando. En realidad, la gloria de la matemática es que *no tenemos que decir de qué estamos hablando*. La gloria es que las leyes, los razonamientos y la lógica son independientes de su "contenido". Si tenemos cualquier otro conjunto de objetos que obedecen el mismo sistema de axiomas que la geometría de Euclides, entonces, si hacemos nuevas definiciones y las utilizamos con una lógica correcta, todas las consecuencias serán correctas, sin importar cuál fue el tema. En la naturaleza, sin embargo, cuando trazamos una línea o establecemos una línea empleando un rayo de luz y un teodolito, como lo hacemos en levantamientos topográficos, ¿estamos midiendo una línea en el sentido de Euclides? No, estamos haciendo una aproximación, los rayos del retículo tienen cierto grosor, pero una línea geométrica no tiene grosor, de manera que si se puede o no emplear la geometría euclidiana para un levantamiento es un asunto físico, no una cuestión matemática. Sin embargo, desde un punto de vista experimental, no desde un punto de vista matemático, necesitamos saber si las leyes de Euclides se aplican a la clase de geometría que utilizamos al medir terrenos. Hacemos una hipótesis de que sea así, y funciona bastante bien, pero no es precisa porque nuestras líneas de levantamiento no son en realidad líneas geométricas. Si esas líneas de Euclides, que en realidad son abstractas, se aplican o no a las líneas de la experiencia es asunto de la experiencia, no es una cuestión que pueda contestarse sólo mediante razonamiento.

Del mismo modo, no podemos decir que $F = ma$ sea una definición, deducir todo de una manera puramente matemática y hacer de la mecánica una teoría matemática, cuando la mecánica es una descripción de la naturaleza. Estableciendo postulados adecuados, siempre es posible construir un sistema matemático, como lo hizo Euclides, pero no podemos hacer una matemática del mundo, porque tarde o temprano tenemos que averiguar si los axiomas son válidos para los objetos de la naturaleza. Así que inmediatamente llegamos a enredarnos con estos complicados y "sucios" objetos de la naturaleza, pero con aproximaciones siempre crecientes en exactitud.

12-2 Fricción

Las consideraciones anteriores demuestran que una verdadera comprensión de las leyes de Newton requiere una discusión de las fuerzas, y el propósito de este capítulo es introducir tal discusión, como una especie de complemento de las leyes de Newton. Ya hemos estudiado las definiciones de aceleración e ideas relacionadas, pero ahora tenemos que estudiar las propiedades de las fuerzas, y este capítulo, a diferencia de los capítulos precedentes, no será muy preciso, porque las fuerzas son bastante complicadas.

Para empezar con una fuerza en particular, consideremos la resistencia al avance que experimenta un aeroplano que vuela en el aire. ¿Cuál es la ley para esa fuerza? (Seguro que existe una ley para cada fuerza, ¡*debemos* tener una ley!) Difícilmente podríamos pensar que la ley para esa fuerza fuera sencilla. Procuren imaginar qué es lo que causa la resistencia al avance de un aeroplano que vuela por el aire —el aire fluyendo velozmente sobre las alas, el remolino en el dorso, los cambios que se originan alrededor del fuselaje y muchas otras complicaciones—, y se darán cuenta de que no habrá una ley sencilla. Por otro lado, es un hecho notable que la fuerza de resistencia sobre un aeroplano sea aproximadamente igual a una constante por el cuadrado de la velocidad, o $F \approx cv^2$.

Ahora, ¿cuál es la naturaleza de esa ley? ¿Es análoga a $F = ma$? De ninguna manera, porque en primer lugar ésta es una ley empírica que se obtiene *grosso modo* por pruebas efectuadas en un túnel de viento. Ustedes dirán: "Bien, puede que $F = ma$ sea empírica también". Ésa no es razón para que exista diferencia. La diferencia no está en que sea empírica, sino que, como entendemos la naturaleza, esta ley es el resultado de un complejo enorme de eventos y no es, fundamentalmente, una cosa sencilla. Si seguimos estudiándola cada vez más y más, midiendo cada vez con mayor exactitud, la ley continuará tornándose *más* complicada y no *menos*. En otras palabras, al estudiar esta ley de resistencia al

avance de un aeroplano más y más detenidamente, encontramos que es "más y más falsa", y mientras más profundamente la estudiamos y mientras más exactamente midamos, tanto más complicada llega a ser la verdad, de manera que en ese sentido consideramos que no resulta de un proceso sencillo y fundamental, lo que concuerda con nuestra suposición original. Por ejemplo, si la velocidad es extremadamente baja, tan baja que un aeroplano común no vuela, como sucede cuando el aeroplano es arrastrado lentamente por el aire, entonces la ley cambia, y más bien la fricción depende linealmente de la velocidad. Para considerar otro ejemplo, la fuerza de fricción sobre una pelota o burbuja, o cualquier cosa que se mueva lentamente en un líquido viscoso como la miel, es proporcional a la velocidad, pero para movimientos tan rápidos que el fluido se arremolina (no sucede con la miel, pero sí con el agua y el aire), entonces la resistencia por fricción es casi proporcional al cuadrado de la velocidad ($F = cv^2$), y si la velocidad continúa aumentando, también esta ley empieza a fallar. Las personas que dicen: "Bien, el coeficiente varía muy poco", están esquivando la discusión. Segundo, existen otras grandes complicaciones: ¿puede descomponerse esta fuerza que actúa sobre el aeroplano o ser analizada como una fuerza que actúa sobre las alas, una fuerza al frente, y así sucesivamente? De hecho, esto puede hacerse, si nos preocupamos de los torques que actúan por aquí y por allá, pero entonces tendremos que obtener leyes especiales para las fuerzas en las alas, etc. Es un hecho asombroso que la fuerza que actúa en un ala dependa de la otra ala; en otras palabras, si desarmamos el aeroplano y colocamos una sola ala en el aire, la fuerza no es la misma que cuando el resto del aeroplano está presente. La razón, por supuesto, es que parte del viento que da al frente se mueve alrededor de las alas y cambia las fuerzas sobre las alas. Parece un milagro que exista tal ley empírica sencilla y aproximada que pueda utilizarse en el diseño de aeroplanos, pero esta ley no se encuentra en la misma clase de las leyes *básicas* de la física, y un estudio más amplio de ella sólo la hará cada vez más complicada. Un estudio sobre cómo depende el coeficiente c de la forma de la parte frontal del aeroplano es, para ponerlo en términos suaves, para desanimar a cualquiera. Simplemente no existe una ley sencilla para determinar el coeficiente en términos de la forma del aeroplano. Por el contrario, la ley de la gravitación es sencilla y un mayor estudio sólo revela una mayor simplicidad.

Acabamos de discutir dos casos de fricción que resultan del movimiento rápido en el aire y el movimiento lento en la miel. Existe otra clase de fricción, llamada fricción seca, o fricción por deslizamiento, que ocurre cuando un cuerpo sólido desliza sobre otro. En este caso se necesita una fuerza para mantener el movimiento. Se la llama fuerza de fricción y su origen también es un asunto muy complicado. Ambas superficies de contacto son irregulares a nivel atómico. Hay muchos puntos de contacto en que los átomos parecen pegarse unos a otros y entonces, al arrastrar el cuerpo que se desliza, los átomos se separan y surge vibración, algo así tiene que suceder. Antes, el mecanismo de esta fricción se creía muy sencilla, que las superficies estaban solamente llenas de irregularidades y la fricción se originaba al deslizar el cuerpo que resbala sobre las protuberancias, pero esto no puede ser, porque no hay pérdida de energía en ese proceso, con lo cual de hecho se consume potencia. El mecanismo de la pérdida de potencia es que al pasar el cuerpo que se desliza por sobre los obstáculos, éstos se deforman y entonces generan ondas y movimientos atómicos y, después de un rato, habrá calor en los dos cuerpos. Ahora bien, es muy notable que de nuevo, empíricamente, esta fricción pueda ser descrita en forma aproximada mediante una ley sencilla. Esta ley postula que la fuerza necesaria para vencer la fricción y arrastrar un objeto sobre otro depende de la fuerza normal (esto es, perpendicular a la superficie) entre las dos superficies en contacto. En realidad, con una buena aproximación, la fuerza de fricción es proporcional a la fuerza normal, y tiene un coeficiente más o menos constante; esto es,

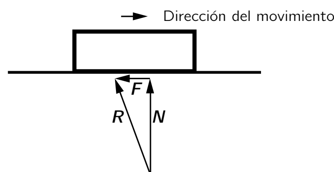


Figura 12-1. La relación entre fuerza de fricción y la fuerza normal para un cuerpo que se desliza.

$$F = \mu N, \tag{12.1}$$

donde μ se llama *coeficiente de fricción* (figura 12-1). A pesar de que este coeficiente no es exactamente constante, la fórmula es una buena regla empírica

para juzgar en forma aproximada la cantidad de fuerza que se necesitaría en ciertas circunstancias prácticas de la ingeniería. Si la fuerza normal o la rapidez del movimiento aumenta demasiado, la ley falla, a causa del excesivo calor generado. Es importante darse cuenta de que cada una de estas leyes empíricas tiene sus limitaciones, más allá de las cuales no dan resultado.

Que la fórmula $F = \mu N$ es aproximadamente correcta puede demostrarse por medio de un sencillo experimento. Armamos un plano inclinado en un ángulo θ pequeño y colocamos un bloque de peso W sobre el plano. Entonces inclinamos el plano para formar un ángulo mayor, hasta que el bloque apenas empieza a moverse a causa de su propio peso. La componente del peso hacia abajo a lo largo del plano es $W \operatorname{sen}\theta$, y ésta debe ser igual a la fuerza de fricción F cuando el bloque se desliza uniformemente. La componente del peso normal al plano es $W\cos\theta$ y ésta es la fuerza normal N. Con estos valores, la fórmula llega a ser $W \operatorname{sen}\theta = \mu W\cos\theta$, de la que obtenemos $\mu = \operatorname{sen}\theta/\cos\theta = \operatorname{tg}\theta$. Si esta ley fuera exactamente válida, un objeto empezaría a deslizarse a una inclinación determinada. Si el mismo bloque se carga con un peso adicional, a pesar de que W aumenta, todas las fuerzas aumentan en la misma proporción y W se cancela. Si μ permaneciera constante, el bloque con el peso adicional se deslizaría otra vez bajo el mismo ángulo de inclinación. Cuando el ángulo θ se determina mediante el experimento con el peso original, se encuentra que con un mayor peso el bloque se deslizará más o menos bajo el mismo ángulo. Esto valdrá aun cuando un cuerpo pese varias veces lo que pesa el otro, y así, concluimos que el coeficiente de fricción es independiente del peso.

Al efectuar este experimento se nota que cuando el plano está inclinado más o menos en el ángulo θ correcto, el bloque no se desliza uniformemente, sino que lo hace en forma vacilante. En algún lugar puede detenerse, en otros puede moverse con cierta aceleración. Este comportamiento indica que el coeficiente de fricción es constante sólo en forma aproximada y varía de lugar en lugar a lo largo del plano. El mismo comportamiento errático se observa estando el bloque cargado o no. Tales variaciones son causadas por diferentes grados de suavidad o rugosidad del plano y tal vez mugre, óxidos u otra materia extraña. Las tablas que dan los supuestos valores de μ para "acero sobre acero", "cobre sobre cobre" y otros por el estilo son todas falsas, porque no toman en cuenta los factores mencionados más arriba, que en realidad son los que determinan μ. La fricción nunca se debe a "cobre sobre cobre", etc., sino que se debe a las impurezas que se adhieren al cobre.

En los experimentos del tipo que se ha descrito anteriormente, la fricción es casi independiente de la velocidad. Muchas personas creen que la fricción que hay que vencer para empezar a mover algo (fricción estática) excede la fuerza necesaria para mantenerlo deslizando (fricción resbalante), pero en metales secos es muy difícil demostrar que existe alguna diferencia. La opinión tiene su origen probablemente en las experiencias en que se presentan pequeñas cantidades de aceite o lubricante, o en que bloques, por ejemplo, son sostenidos por resortes u otros soportes flexibles tal que parecen estar ligados.

Es bastante difícil efectuar experimentos cuantitativos exactos sobre la fricción, y las leyes de la fricción todavía no han sido muy bien analizadas, a pesar del alto valor que tiene para la ingeniería un análisis exacto. A pesar de que la ley $F = \mu N$ es bastante exacta una vez que las superficies han sido estandarizadas, la razón para esta forma de ley en realidad no se comprende. Demostrar que μ es casi independiente de la velocidad exige cierta experimentación delicada, porque la fricción aparente disminuye mucho si la superficie inferior vibra muy rápido. Cuando el experimento se realiza a velocidades muy altas, hay que tener cuidado en que los objetos no vibren uno respecto del otro, ya que las disminuciones aparentes de la fricción a altas velocidades se deben a menudo a vibraciones. De todos modos, esta ley de la fricción es otra de esas leyes casi empíricas que no se comprenden perfectamente, y en vista de todo el trabajo que se ha hecho, es sorprendente que no se haya llegado a una mayor comprensión de este fenómeno. Actualmente, es imposible aún estimar el coeficiente de fricción entre dos sustancias.

Ha quedado indicado más arriba que ensayos para medir μ haciendo resbalar sustancias puras, como cobre sobre cobre, llevan a resultados falsos, porque las superficies en contacto no son de cobre puro, sino mezclas de óxido y otras impurezas. Si procuramos obtener cobre absolutamente puro, si limpiamos y pulimos las superficies, desgastamos los materiales al vacío y tomamos todas las precauciones posibles, todavía no se obtiene μ. Porque, si inclinamos el aparato aun hasta una posición vertical, el cuerpo que desliza no va a caer —¡los dos pedazos de cobre se adhieren!—. ¡El coeficiente μ, que generalmente es menor que la unidad para superficies razonablemente duras, llega a ser varias veces la unidad! La razón de este comportamiento inesperado es que cuando todos los átomos en contacto son de la misma naturaleza, no hay manera de que los átomos "sepan" que están en diferentes pedazos de cobre. Cuando hay otros átomos, en los óxidos y en las grasas y láminas superficiales delgadas más complicadas de contaminantes entre ellas, los átomos "saben" cuándo no están en la misma parte. Cuando consideramos que son las fuerzas entre los átomos las que mantienen unido el cobre como un sólido, debe quedar claro que es imposible obtener el coeficiente correcto de fricción para metales puros.

El mismo fenómeno puede observarse en un sencillo experimento casero, ejecutado con una placa de vidrio plano y un vaso de vidrio. Si se coloca el vaso sobre la placa y se tira con un hilo desliza bastante bien y uno puede sentir el coeficiente de fricción, es un poquito irregular, pero es un coeficiente. Si ahora mojamos la placa y la parte inferior del vaso y tiramos de él otra vez, encontramos que se adhieren y si observamos detenidamente, encontraremos que se rayan, porque el agua puede separar la grasa y otras materias extrañas de la placa, y entonces tenemos en realidad contacto entre vidrio y vidrio; este contacto es tan bueno que se mantiene ajustado y se resiste tanto a la separación que el vidrio se rompe, es decir, produce rayas.

12-3 Fuerzas moleculares

A continuación discutiremos las características de las fuerzas moleculares. Éstas son fuerzas entre los átomos y constituyen el origen último de la fricción. Las fuerzas moleculares nunca han sido explicadas satisfactoriamente sobre la base de la física clásica, se necesita de la mecánica cuántica para entenderlas plenamente. Sin embargo, empíricamente, la fuerza entre átomos es como se ilustra esquemáticamente en la figura 12-2, en que se grafica la fuerza entre átomos en función de la distancia r entre ellos. Existen diferentes casos: en la molécula de agua, por ejemplo, las cargas negativas se encuentran en mayor cantidad en el oxígeno y las posiciones medias de las cargas negativas y de las cargas positivas no se encuentran en el mismo punto; en consecuencia, otra molécula cercana experimenta una fuerza relativamente grande que se denomina fuerza dipolo-dipolo. Sin embargo, para muchos sistemas las cargas están mucho mejor compensadas, en particular en el oxígeno gaseoso, que es perfectamente simétrico. En este caso, a pesar de que las cargas negativas y positivas están dispersas sobre la molécula, la distribución es tal que el centro de las cargas negativas y el centro de las cargas positivas coinciden. Una molécula en la cual no coinciden los centros se denomina molécula polar, y la carga multiplicada por la separación de los centros se llama momento dipolar. Una molécula no polar es aquella en que los centros de las cargas coinciden. Para todas las moléculas no polares, en que todas las fuerzas eléctricas se neutralizan, de todos modos resulta que la fuerza a una distancia muy grande es una atracción y varía inversamente con la potencia elevada a la siete de la distancia, o $F = k/r^7$, en donde k es una constante que depende de las moléculas. Por qué es así, lo sabremos sólo cuando aprendamos mecánica cuántica. Cuando hay dipolos las fuerzas son más grandes. Cuando los átomos o las moléculas se acercan demasiado se repelen con una fuerza repulsiva muy grande, ¡esto es lo que impide que al caer atravesemos el piso!

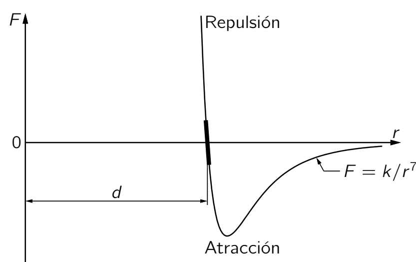


Figura 12-2. La fuerza entre dos átomos en función de su distancia de separación.

Estas fuerzas moleculares pueden mostrarse de una manera bastante directa; una de éstas es el experimento de fricción con un vaso de vidrio que se desliza,

otro consiste en tomar dos superficies cuidadosamente pulidas y limpias que son exactamente planas, de manera que las superficies puedan juntarse muy estrechamente. Un ejemplo de estas superficies son los bloques Johansson que se emplean en talleres de máquinas como patrones para medir longitudes con precisión. Si uno de estos bloques se hace deslizar sobre otro muy cuidadosamente y el de arriba se levanta, el otro se adherirá y también se levantará debido a las fuerzas moleculares, constituyendo un ejemplo de atracción directa entre los átomos de un bloque por los átomos del otro.

Sin embargo, estas fuerzas moleculares de atracción no son todavía fundamentales, en el sentido en que la gravitación es fundamental, se deben a la vasta y compleja interacción de todos los electrones y núcleos de una molécula con todos los electrones y núcleos de la otra. Cualquier fórmula de apariencia sencilla que obtengamos representa una suma de complicaciones, de manera que todavía no hemos obtenido el fenómeno fundamental.

Como las fuerzas moleculares atraen a grandes distancias y repelen a distancias cortas, como se muestra en la figura 12-2, podemos formar sólidos en que todos los átomos se mantienen unidos por sus atracciones, y se mantienen separados por la repulsión que se origina cuando están demasiado juntos. A una cierta distancia d (donde la gráfica de la figura 12-2 cruza el eje) las fuerzas son cero, lo que significa que están equilibradas, de manera que las moléculas se mantienen separadas a esa distancia entre sí. Si las moléculas se comprimen hasta acercarse a una distancia inferior a d, todas muestran una repulsión representada por la parte de la gráfica que se encuentra sobre el eje r. Comprimir las moléculas sólo ligeramente exige una gran fuerza, porque la repulsión molecular rápidamente se devuelve muy grande a distancias inferiores a d. Si las moléculas se separan ligeramente, se presenta una leve atracción que aumenta al incrementar la separación. Si son separadas con fuerza suficiente, se distanciarán permanentemente —la ligadura se rompe—.

Si las moléculas son comprimidas sólo una distancia *muy pequeña*, o separadas sólo una distancia *muy pequeña* más allá de d, la correspondiente distancia a lo largo de la curva de la figura 12-2 es también muy pequeña y puede aproximarse a una línea recta. Por consiguiente, en muchas circunstancias, si el desplazamiento no es demasiado grande, *la fuerza es proporcional al desplazamiento*. Este principio se conoce como ley de Hooke, o ley de la elasticidad, que dice que la fuerza que actúa sobre un cuerpo y que trata de restaurar el cuerpo a su condición original cuando ha sido deformado es proporcional a la deformación. Esta ley, por supuesto, es válida sólo cuando la deformación es relativamente pequeña, cuando aumenta demasiado, el cuerpo se despedaza o quiebra dependiendo de la clase de deformación. Los valores de la fuerza, para los cuales la ley de Hooke es válida, dependen del material. Por ejemplo, para masa o masilla la fuerza es muy pequeña, pero para el acero es relativamente grande. La ley de Hooke puede ser demostrada bastante bien por medio de un resorte largo hecho de acero y suspendido verticalmente. Un peso apropiado colgado en el extremo inferior del resorte produce una pequeña torsión a lo largo del alambre, que resulta en una pequeña desviación vertical en cada vuelta y suma un gran desplazamiento si tiene muchas vueltas. Si se mide el alargamiento total producido, digamos que por un peso de 100 gramos, se encuentra que agregando pesas de 100 gramos, cada una producirá un alargamiento adicional que es muy cercanamente igual al estiramiento que se midió para los primeros 100 gramos. Este cociente constante entre la fuerza y el alargamiento empieza a cambiar cuando se sobrecarga el resorte, esto es, la ley de Hooke ya no se cumple.

12-4 Fuerzas fundamentales. Campos

Ahora discutiremos las únicas fuerzas restantes que son fundamentales. Las llamamos fundamentales en el sentido de que sus leyes son fundamentalmente sencillas. Primero discutiremos la fuerza eléctrica. Los objetos son portadores de cargas eléctricas que simplemente consisten en electrones o protones. Si dos cuerpos cualesquiera están cargados eléctricamente, existe una fuerza eléctrica

entre ellos y si las magnitudes de las cargas son q y q_2, respectivamente, la fuerza varía en razón inversa al cuadrado de la distancia entre las cargas o bien $F = (\text{const.})q_1q_2/r^2$. Para cargas distintas, esta ley es similar a la ley de gravitación, pero para cargas de signos *iguales* la fuerza es repulsiva y el signo (dirección) se invierte. Las cargas q_1 y q_2 pueden ser intrínsecamente positivas o negativas y en cualquier aplicación específica de la fórmula, la dirección de la fuerza resultará correcta si a las q se les asigna el signo más o menos apropiado, la fuerza está dirigida a lo largo de la línea entre las dos cargas. En la fórmula dicha constante depende, por supuesto, de las unidades que se usen para la fuerza, la carga y la distancia. En la práctica la carga se mide en coulomb, la distancia en metros y la fuerza en newton. Entonces, para obtener la fuerza en newton, la constante (que por razones históricas se escribe $1/4\pi\epsilon_0$) toma el valor numérico:

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ coulomb}^2/\text{newton} \cdot \text{m}^2$$

o

$$1/4\pi\epsilon_0 = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2.$$

De manera que la ley de la fuerza para cargas estáticas es:

$$\boldsymbol{F} = q_1q_2\boldsymbol{r}/4\pi\epsilon_0r^3. \tag{12.2}$$

En la naturaleza, la carga más importante de todas es la de un electrón, que es $1.60 = 10^{-19}$ coulomb. Al trabajar con fuerzas eléctricas entre partículas fundamentales en vez de cargas grandes, muchos prefieren la combinación $(q_{\text{el}})^2/4\pi\epsilon_0$, en que q_{el} se define como la carga de un electrón. Esta combinación ocurre frecuentemente y para simplificar los cálculos se ha definido por el símbolo e^2; su valor numérico en el sistema mks de unidades resulta ser $(1.52 \times 10^{-14})^2$. La ventaja de usar la constante en esta forma es que la fuerza entre dos electrones, en newtons, puede escribirse sencillamente como e^2/r^2, con r en metros, sin ninguna de las constantes individuales. Las fuerzas eléctricas son mucho más complicadas que lo que indica esta sencilla fórmula, ya que la fórmula asigna la fuerza entre dos objetos sólo cuando los objetos están en reposo. Pronto consideraremos el caso más general.

En el análisis de fuerzas de naturaleza más fundamental (no de fuerzas como la fricción, sino de fuerzas eléctricas o gravitacionales), se ha desarrollado un concepto muy interesante e importante. Ya que a primera vista las fuerzas son mucho más complicadas que lo que indica la ley de la inversa del cuadrado y estas leyes son válidas sólo cuando los cuerpos que interactúan están en reposo, se necesita un método mejor para tratar las fuerzas muy complejas que resultan cuando los cuerpos empiezan a moverse en forma complicada. La experiencia ha demostrado que un enfoque, conocido como concepto de “campo”, es muy útil en el análisis de fuerzas de este tipo. Con el fin de ilustrar la idea para fuerzas eléctricas, digamos, supongan que tenemos dos cargas eléctricas q_1 y q_2 colocadas en los puntos P y R, respectivamente. Entonces la fuerza entre las cargas está dada por:

$$\boldsymbol{F} = q_1q_2\boldsymbol{r}/4\pi\epsilon_0r^3. \tag{12.3}$$

Para analizar esta fuerza mediante el concepto de campo, decimos que la carga q_1 en P produce una “condición” en R, tal que cuando la carga q_2 se coloca en R “siente” la fuerza. Ésta es una forma, extraña quizá, de describirla: decimos que la fuerza $\boldsymbol{F}$ sobre q_2 en R puede ser escrita en dos partes. Es q_2 multiplicada por una cantidad $\boldsymbol{E}$ que estaría ahí, estuviese o no q_2 (siempre que mantengamos todas las demás cargas en sus respectivos lugares). Decimos que $\boldsymbol{E}$ es la “condición” producida por q_1 y $\boldsymbol{F}$ es la respuesta de q_2 a $\boldsymbol{E}$. $\boldsymbol{E}$ se denomina *campo eléctrico* y es un vector. La fórmula para el campo eléctrico $\boldsymbol{E}$ que se produce en el punto R debido a la carga q_1 en P es la carga q_1 multiplicada por la constante $1/4\pi\epsilon_0$ dividido entre r^2 (r es la distancia de P a R) y actúa en la dirección del radio vector (el radio vector $\boldsymbol{r}$ dividido entre su propia longitud). La expresión para $\boldsymbol{E}$

es así:

$$\boldsymbol{E} = q_1\boldsymbol{r}/4\pi\epsilon_0 r^3. \tag{12.4}$$

Entonces escribimos

$$\boldsymbol{F} = q_2\,\boldsymbol{E}, \tag{12.5}$$

que expresa la fuerza, el campo y la carga en el campo. ¿Cuál es el propósito de todo esto? El propósito es dividir el análisis en dos partes. Una parte dice que algo *produce* un campo. La otra parte dice que algo es *influenciado* por el campo. Al observar estas dos partes independientemente, esta separación en el análisis simplifica el cálculo de un problema en muchas situaciones. Si están presentes muchas cargas, primero calculamos el campo eléctrico total producido en R por todas las cargas, y luego, conociendo la carga que se coloca en R, encontramos la fuerza que actúa sobre ella.

En el caso de la gravitación podemos hacer exactamente lo mismo. En este caso, en que la fuerza $\boldsymbol{F} = -Gm_1m_2\boldsymbol{r}/r^3$, podemos hacer un análisis análogo, como sigue: la fuerza que actúa sobre un cuerpo en un campo gravitacional es igual a la masa del cuerpo por el campo $\boldsymbol{C}$. La fuerza sobre m_2 es la masa m_2 multiplicada por el campo $\boldsymbol{C}$ producido por m_1, esto es, $\boldsymbol{F} = m_2\,\boldsymbol{C}$. Entonces el campo $\boldsymbol{C}$ producido por un cuerpo de masa m_1 es $\boldsymbol{C} = -Gm_1\boldsymbol{r}/r^3$ y está dirigido radialmente como en el caso eléctrico.

A pesar de lo que pueda parecer en un principio, separar una parte de la otra no es una trivialidad. Sería trivial, sólo otra manera de escribir lo mismo, si las leyes de fuerza fueran sencillas, pero las leyes de fuerza son tan complicadas, que resulta que los campos tienen una realidad que es casi independiente de los objetos que los crean. Uno puede hacer algo así como agitar una carga y producir un efecto, un campo, a cierta distancia, entonces si uno deja de agitar la carga, el campo sigue el curso de todo lo pasado, porque la interacción de las dos partículas no es instantánea. Es deseable tener alguna forma de recordar lo que pasó previamente. Si la fuerza sobre alguna carga depende de donde estaba otra carga ayer, y sí es así, entonces necesitamos un mecanismo para seguir la pista de lo que sucedió ayer, y ésta es la naturaleza del campo. De manera que, cuando las fuerzas se vuelven más complicadas, el campo se hace más y más real y esta técnica llega a ser cada vez menos una separación artificial.

Al analizar las fuerzas por medio de campos, necesitamos dos clases de leyes relacionadas con los campos. La primera es la respuesta a un campo, que da las ecuaciones de movimiento. Por ejemplo, la ley de respuesta de una masa a un campo gravitacional es que la fuerza es igual a la masa por el campo gravitacional, o bien, si también existe una carga en el cuerpo, la respuesta de la carga al campo eléctrico es igual a la carga por el campo eléctrico. La segunda parte del análisis de la naturaleza en estas situaciones es formular las leyes que determinan la intensidad del campo y cómo se produce. A veces a estas leyes se las denomina *ecuaciones de campo*. Aprenderemos más sobre ellas a su debido tiempo, pero escribiremos algunas cosas ahora.

Primero, el hecho más notable de todos, el cual es exactamente verdadero y que se puede comprender fácilmente, es que el eléctrico total producido por varias fuentes es la suma vectorial de los campos eléctricos producidos por la primera fuente, la segunda y así sucesivamente. En otras palabras, si tenemos numerosas cargas produciendo un campo y si una sola puede producir el campo $\boldsymbol{E}_1$, otra puede producir el campo $\boldsymbol{E}_2$ y así sucesivamente, entonces sumamos los vectores para obtener el campo total. Este principio puede expresarse así:

$$\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2 + \boldsymbol{E}_3 + \cdots \tag{12.6}$$

o, en vista de las definiciones dadas más arriba,

$$\boldsymbol{E} = \sum_i \frac{q_i\boldsymbol{r}_i}{4\pi\epsilon_0 r_i^3}. \tag{12.7}$$

¿Puede aplicarse el mismo método a la gravitación? La fuerza entre dos masas m_1 y m_2 fue expresada por Newton como $\boldsymbol{F} = -Gm_1m_2\boldsymbol{r}/r^3$. Pero según el

concepto de campo, podemos decir que m_1 crea un campo $\boldsymbol{C}$ en todo el espacio que la rodea, de manera que la fuerza sobre m_2 está dada por:

$$\boldsymbol{F} = m_2\boldsymbol{C}. \tag{12.8}$$

Por analogía completa con el caso eléctrico

$$\boldsymbol{C}_i = -Gm_i\boldsymbol{r}_i/r_i^3 \tag{12.9}$$

y el campo gravitacional producido por varias masas es:

$$\boldsymbol{C} = \boldsymbol{C}_1 + \boldsymbol{C}_2 + \boldsymbol{C}_3 + \cdots \tag{12.10}$$

En el capítulo 9, al resolver un caso sobre movimiento planetario en esencia, empleamos este principio. Sencillamente sumamos todos los vectores fuerza para obtener la fuerza resultante sobre un planeta. Si dividimos entre la masa del planeta en cuestión, obtenemos la ecuación (12.10).

Las ecuaciones (12.6) y (12.10) expresan lo que se conoce como *principio de superposición* de campos. Este principio dice que el campo total debido a todas las fuentes es la suma de los campos debidos a cada fuente. Por lo que sabemos, para la electricidad ésta es una ley que podemos garantizar de una manera absoluta, que es válida aunque la ley de fuerza es complicada a causa de los movimientos de las cargas. Hay violaciones aparentes, pero un análisis más cuidadoso ha demostrado siempre que se deben a no haber tomado en cuenta cierta carga en movimiento. Sin embargo, a pesar de que el principio de superposición se aplica con exactitud a fuerzas eléctricas, no es exacto para la gravedad si el campo es demasiado intenso y la ecuación de Newton (12.10) es sólo aproximada según la teoría gravitacional de Einstein.

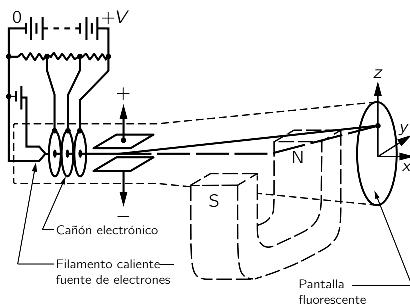


Figura 12-3. Un tubo de rayos de electrones.

Estrechamente relacionada con la fuerza eléctrica existe otra clase, denominada fuerza magnética, y ésta también se analiza en términos de un campo. Algunas de las relaciones cualitativas entre fuerzas eléctricas y magnéticas pueden ser ilustradas mediante un experimento con un tubo de rayos electrónicos (figura 12-3). En un extremo de ese tubo hay una fuente que emite un chorro de electrones. Dentro del tubo hay dispositivos para acelerar los electrones a una velocidad alta y enviar algunos de ellos en un haz estrecho a una pantalla fluorescente en el otro extremo del tubo. Un punto luminoso brilla en el centro de la pantalla donde golpean los electrones y esto hace posible trazar la trayectoria de los electrones. En el trayecto hacia la pantalla, el haz de electrones pasa por un espacio estrecho entre un par de placas metálicas paralelas colocadas, digamos, horizontalmente. Se puede aplicar un voltaje entre las placas metálicas, de manera que cualquier placa puede hacerse negativa a voluntad. Cuando hay un voltaje así, existe un campo eléctrico entre las placas.

La primera parte del experimento consiste en suministrar un voltaje negativo a la placa inferior, lo cual significa que se han colocado electrones adicionales en la placa inferior. Como cargas iguales se repelen, el punto luminoso en la pantalla se desplaza instantáneamente hacia arriba. (También podríamos decir esto de otra manera, los electrones "sintieron" el campo y respondieron desviándose hacia arriba.) A continuación invertimos el voltaje haciendo negativa la placa *superior*. El punto luminoso en la pantalla salta ahora debajo del centro, demostrando que los electrones del haz fueron repelidos por la placa de encima de ellos (o bien, podríamos decir otra vez que los electrones "respondieron" al campo, que ahora tiene dirección inversa).

La segunda parte del experimento consiste en desconectar el voltaje de las placas y probar el efecto de un campo magnético sobre el haz de electrones. Esto se hace por medio de un imán en forma de herradura, cuyos polos están lo suficientemente separados como para abarcar más o menos el tubo. Supongan que sujetamos el imán por debajo del tubo con la misma orientación de la letra U, con sus polos hacia arriba y con una parte del tubo por el medio. Notamos que el punto luminoso se desvía, digamos, hacia arriba, al acercar el imán al tubo desde abajo. De manera que parece que el imán repele el haz de electrones.

Sin embargo, esto no es tan sencillo, porque si invertimos el imán sin invertir los polos lado por lado y acercamos al tubo desde arriba, el punto luminoso se mueve todavía *hacia arriba*, de manera que el haz de electrones *no* es repelido; por el contrario, parece que fuera atraído esta vez. Ahora, empecemos otra vez, volviendo el imán a su orientación U original y manteniéndolo por debajo del tubo como antes. Sí, el punto todavía se desvía hacia arriba, pero ahora giremos el imán en 180° alrededor de un eje vertical, de manera que todavía esté en la posición U, pero los polos están invertidos lado por lado. Observen: ahora el punto salta hacia abajo y se mantiene abajo, aunque invirtamos el imán y lo acerquemos desde arriba como antes.

Para comprender este extraño comportamiento, debemos tener una nueva combinación de fuerzas. Lo explicamos: entre los polos del imán existe un *campo magnético.* Este campo tiene una dirección que siempre sale de un polo en particular (que podríamos marcar) y entra al otro. Al invertir el imán no cambió la dirección del campo, pero al invertir los polos lado por lado se invirtió la dirección. Por ejemplo, si la velocidad de los electrones fuera horizontal en la dirección x y el campo magnético fuera también horizontal, pero en la dirección y, la fuerza magnética *sobre los electrones en movimiento* sería en la dirección z, esto es, hacia arriba o hacia abajo dependiendo de si el campo estaba en la dirección y positiva o negativa.

A pesar de que ahora no daremos la ley correcta de la fuerza entre las cargas que se mueven en forma arbitraria, unas respecto a otras, porque es demasiado complicada, daremos un aspecto de ella, la ley completa de las fuerzas, *si se conocieran los campos.* La fuerza que actúa sobre un objeto cargado depende de su movimiento, si existiese una fuerza, cuando el objeto está en reposo en un lugar determinado, ésta se torna proporcional a la carga, el coeficiente es lo que llamamos *campo eléctrico.* Cuando el objeto se mueve, la fuerza puede ser diferente y la corrección, la nueva "parte" de la fuerza, resulta ser exactamente dependiente *de la velocidad de manera lineal*, pero *perpendicular* a $\boldsymbol{v}$ y a otra cantidad vectorial que llamamos *inducción magnética* $\boldsymbol{B}$. Si las componentes del campo eléctrico $\boldsymbol{E}$ y la inducción magnética $\boldsymbol{B}$ son, respectivamente, (E_x, E_y, E_z) y (B_x, B_y, B_z), y si la velocidad $\boldsymbol{v}$ tiene las componentes $(v_x,\, v_y,\, v_z)$, entonces la fuerza magnética y eléctrica total que actúa sobre una carga q en movimiento tiene las componentes

$$\begin{aligned} F_x &= q(E_x + v_y B_z - v_z B_y), \\ F_y &= q(E_y + v_z B_x - v_x B_z), \\ F_z &= q(E_z + v_x B_y - v_y B_x). \end{aligned} \tag{12.11}$$

Si, por ejemplo, la única componente del campo magnético fuera B_y y la única componente de la velocidad fuera v_x, el único término que queda en la fuerza magnética sería una fuerza en la dirección z, perpendicular a $\boldsymbol{B}$ y a $\boldsymbol{v}$.

12-5 Pseudofuerzas

La siguiente clase de fuerza que discutiremos podría llamarse pseudofuerza. En el capítulo 11 discutimos la relación entre dos personas, Juan y Pedro, que utilizan distintos sistemas de coordenadas. Supongamos que las posiciones de una partícula mediadas por Juan sean x y por Pedro x', entonces las leyes son como sigue:

$$x = x' + s, \qquad y = y', \qquad z = z',$$

donde s es el desplazamiento del sistema de Pedro relativo al de Juan. Si suponemos que las leyes de movimiento son correctas para Juan, ¿qué aspecto tienen para Pedro? Primero encontramos que:

$$dx/dt = dx'/dt + ds/dt.$$

Anteriormente consideramos el caso en que s era constante y encontramos que s no influía en las leyes de movimiento, ya que $ds/dt = 0$; en última instancia,

entonces, las leyes de la física eran las mismas en ambos sistemas. Pero otro caso que podemos considerar es que $s = ut$, en que u es una velocidad uniforme en línea recta. Entonces s no es constante y ds/dt no es cero, sino que es u, una constante. Sin embargo, la aceleración d^2x/dt^2 es aún igual a d^2x'/dt^2, porque $du/dt = 0$. Esto demuestra la ley que empleamos en el capítulo 10, esto es, que si nos movemos en línea recta con velocidad uniforme, las leyes de la física nos parecerán las mismas que cuando estamos en reposo. Ésta es la transformación de Galileo. Pero queremos discutir el interesante caso en que s es todavía más complicado, digamos $s = at^2/2$. Entonces $ds/dt = at$ y $d^2s/dt^2 = a$, una aceleración uniforme, o bien, en un caso todavía más complicado, la aceleración podría ser función del tiempo. Esto significa que a pesar de que las leyes de fuerza desde el punto de vista de Juan serían

$$m\,\frac{d^2x}{dt^2} = F_x,$$

las leyes de la fuerza como vistas por Pedro parecerían

$$m\,\frac{d^2x'}{dt^2} = F_{x'} = F_x - ma.$$

Esto es, como el sistema de coordenadas de Pedro está acelerando respecto al de Juan, el término extra ma entra en escena, y Pedro tendrá que corregir sus fuerzas en esa cantidad para conseguir que funcionen las leyes de Newton. En otras palabras, he aquí una nueva fuerza aparente y misteriosa de origen desconocido, que aparece, por supuesto, porque Pedro tiene un sistema de coordenadas equivocado. Éste es un ejemplo de pseudofuerza, otros ejemplos ocurren en sistemas de coordenadas que *rotan*.

Otro ejemplo de pseudofuerza es la que a menudo se denomina "fuerza centífruga". Un observador en un sistema de coordenadas que rota, por ejemplo una caja que rota, encontrará fuerzas misteriosas no explicadas por ningún origen de fuerza conocido, que lanza los objetos hacia afuera contra las paredes. Estas fuerzas se deben sencillamente al hecho de que el observador no tiene el sistema de coordenadas de Newton, el cual es el sistema de coordenadas más sencillo.

Una pseudofuerza puede ilustrarse por medio de un interesante experimento en que empujamos un jarro con agua sobre una mesa con aceleración. La gravedad, por supuesto, actúa hacia abajo sobre el agua, pero debido a la aceleración horizontal existe también una pseudofuerza que actúa horizontalmente y en sentido contrario a la aceleración. La resultante de la gravedad y la pseudofuerza forma un ángulo con la vertical y durante la aceleración la superficie del agua será perpendicular a la fuerza resultante, esto es, oblicua respecto a la mesa, con el agua a un nivel más alto en el lado posterior del jarro. Cuando deja de actuar la fuerza que empuja al jarro y disminuye la velocidad por la fricción, la pseudofuerza se invierte, y el nivel del agua es más alto en la parte delantera del jarro (figura 12-4).

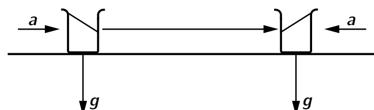


Figura 12-4. Ejemplo de pseudofuerza.

Un aspecto muy importante de las pseudofuerzas es que son siempre proporcionales a las masas, lo mismo es cierto para la gravedad. Existe la posibilidad, por consiguiente, que la *gravedad misma sea una pseudofuerza*. ¿No es posible, tal vez, que la gravitación se deba simplemente al hecho de que no tenemos el sistema de coordenadas apropiado? Después de todo, siempre podemos obtener una fuerza proporcional a la masa si imaginamos que un cuerpo está acelerando. Por ejemplo, un hombre encerrado en una caja en reposo sobre la tierra se encuentra sujeto al piso del cajón con una cierta fuerza que es proporcional a su masa. Pero si no estuviera la tierra y el cajón estuviese en reposo, el hombre dentro del cajón flotaría en el espacio. Por otro lado, si no existiera la tierra y algo estuviera *tirando* del cajón proporcionándole la aceleración g, el hombre en el cajón, analizando la física, encontraría una pseudofuerza que lo tiraría al piso, como lo hace la gravedad.

Einstein propuso la famosa hipótesis de que las aceleraciones imitan a la gravitación, que las fuerzas de aceleración (las pseudofuerzas) *no pueden distinguirse*

de las de gravedad, no es posible afirmar qué parte de una fuerza es gravedad y qué parte es pseudofuerza.

Puede que esté bien considerar la gravedad como una pseudofuerza, decir que todos estamos sujetos a la tierra, porque estamos acelerando hacia arriba, pero qué diremos de la gente que vive en Madagascar, en el otro lado de la tierra, ¿aceleran también ellos? Einstein encontró que la gravedad podría considerarse como una pseudofuerza sólo en un punto a la vez, y sus consideraciones lo condujeron a sugerir que la *geometría del mundo* es más complicada que la geometría euclidiana común. La presente discusión es sólo cualitativa y no pretende dar una idea general. Para dar una idea aproximada de cómo la gravitación podría ser el resultado de pseudofuerzas, presentamos un ejemplo que es solamente geométrico y no representa la situación real. Supongan que vivimos en dos dimensiones y que no sabemos nada de una tercera. Pensamos que nos encontramos en un plano, pero supongan que estamos realmente en la superficie de una esfera. Además, supongan que disparamos un objeto sobre este terreno, sin que actúe alguna fuerza sobre él. ¿A dónde irá? Parecería ir en una línea recta, pero tienen que permanecer en la superficie de la esfera, en que la distancia más corta entre dos puntos es a lo largo de un círculo máximo de manera que se mueve en un círculo máximo. Si disparamos en igual forma otro objeto, pero en otra dirección, avanza en otro círculo máximo. Porque pensamos que estamos en un plano, esperamos que los dos cuerpos continúen apartándose linealmente con el tiempo, pero una observación cuidadosa demostrará que si avanzan suficientemente empiezan a juntarse otra vez, como si estuvieran atrayéndose. Pero *no* se están atrayendo —sólo existe algo "misterioso" en esta geometría—. Este ejemplo en particular no describe correctamente la forma en que la geometría de Einstein es "misteriosa", pero ilustra que si deformamos suficientemente la geometría es posible que toda gravitación esté relacionada en alguna forma con pseudofuerzas; ésa es la idea general de la teoría einsteniana de la gravitación.

12-6 Fuerzas nucleares

Finalizamos este capítulo con una breve discusión sobre las únicas fuerzas restantes, que se denominan *fuerzas nucleares*. Estas fuerzas se encuentran dentro de los núcleos de los átomos y, a pesar de que han sido muy discutidas, nadie ha calculado jamás la fuerza que actúa entre dos núcleos y, en realidad, hoy día no se conoce ninguna ley para las fuerzas nucleares. Estas fuerzas tienen un alcance muy pequeño, que es más o menos como el tamaño del núcleo, tal vez 10^{-13}cm. Con partículas tan pequeñas y a distancias tan cortas, sólo son válidas las leyes de la mecánica cuántica, no así las leyes newtonianas. En el análisis nuclear ya no pensamos en función de fuerzas y, en realidad, podemos remplazar el concepto de fuerza por el de energía de interacción entre dos partículas, un tema que se discutirá más tarde. Cualquier fórmula que se pueda escribir sobre fuerzas nucleares es en cierto modo una burda aproximación que omite muchas complicaciones, una puede ser como la que sigue: las fuerzas dentro del núcleo no varían en razón inversa al cuadrado de la distancia, sino que se extinguen exponencialmente en cierta distancia r como se expresa en $F = (1/r^2)\exp(-r/r_0)$, donde la distancia r_0 es del orden de 10^{-13} centímetros. En otras palabras, las fuerzas desaparecen tan pronto como las partículas se encuentren a una distancia mayor que ésa a pesar de que son fuerzas muy grandes dentro del alcance 10^{-13} centímetros. Como se entienden hoy día, las leyes de las fuerzas nucleares son muy complejas, no las entendemos de una manera sencilla y el problema de analizar el mecanismo fundamental de las fuerzas nucleares no está resuelto. Intentos de una solución han llevado al descubrimiento de numerosas partículas extrañas, los mesones π, por ejemplo, pero el origen de estas fuerzas permanece oscuro.

13

Trabajo y energía potencial (A)

13-1 Energía de un cuerpo que cae

En el capítulo 4 discutimos la conservación de la energía. En esa discusión no empleamos las leyes de Newton, pero es, por supuesto, de gran interés entender cómo sucede que esa energía de hecho se conserve de acuerdo con estas leyes. Para mayor claridad empezaremos con el ejemplo más simple posible, y después desarrollaremos ejemplos más y más difíciles.

El ejemplo más simple de conservación de energía lo constituye un objeto que cae verticalmente, que se mueve sólo en dirección vertical. Un objeto que cambia su altura bajo la influencia de la gravedad solamente, tiene una energía cinética T (o EC) debida a su movimiento durante la caída, y una energía potencial mgh, abreviada U (o EP), cuya suma es constante:

$$\underset{\text{EC}}{\tfrac{1}{2}mv^2} + \underset{\text{EP}}{mgh} = \text{constante},$$

o

$$T + U = \text{constante}. \tag{13.1}$$

Ahora nos gustaría demostrar la validez de esta afirmación. ¿Qué queremos decir con demostrar su validez? A partir de la segunda ley de Newton podemos formular fácilmente cómo se mueve el objeto, y es fácil averiguar que la velocidad cambia con el tiempo, es decir, que aumenta en proporción al tiempo y que la altura varía con el cuadrado del tiempo. De manera que si medimos la altura a partir de un punto cero donde el objeto está en reposo, no es ningún milagro que la altura sea igual al cuadrado de la velocidad multiplicada por un número de constantes. Sin embargo, observamos esto un poco más detenidamente.

Averigüemos *directamente* a partir de la segunda ley de Newton cómo debe variar la energía cinética, tomando la derivada de la energía cinética respecto al tiempo y después empleando las leyes de Newton. Derivando $\frac{1}{2}mv^2$ en relación al tiempo, obtenemos

$$\frac{dT}{dt} = \frac{d}{dt}\left(\tfrac{1}{2}mv^2\right) = \tfrac{1}{2}m2v\,\frac{dv}{dt} = mv\,\frac{dv}{dt}, \tag{13.2}$$

ya que m se supone constante. Pero según la segunda ley de Newton, $m(dv/dt) = F$, de modo que

$$dT/dt = Fv. \tag{13.3}$$

En general, resultará ser $\boldsymbol{F} \cdot \boldsymbol{v}$, pero en nuestro caso unidimensional dejémoslo como fuerza por velocidad.

Ahora bien, en nuestro ejemplo simple la fuerza es constante, igual a $-mg$, una fuerza vertical (el signo menos indica que actúa hacia abajo), y la velocidad, por supuesto, es la derivada de la posición vertical, o altura h, respecto al tiempo. Así la derivada de la energía cinética es $-mg(dh/dt)$, cantidad que, milagro de milagros, ¡es menos la derivada de otra cosa! ¡Es menos la derivada respecto al tiempo de mgh! Por consiguiente, a medida que pasa el tiempo, las variaciones de energía cinética y de la cantidad mgh son iguales y opuestas, de manera que la suma de las dos cantidades permanece constante. Q.E.D.

Hemos demostrado, partiendo de la segunda ley de Newton del movimiento, que la energía se conserva en los casos de fuerzas constantes, cuando sumamos la energía potencial mgh a la energía cinética $\frac{1}{2}mv^2$. Ahora estudiemos esto un poco más y veamos si se puede generalizar y así avanzar en nuestro conocimiento.

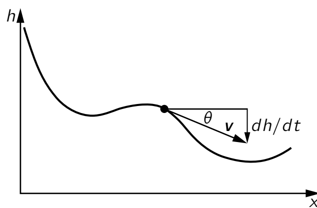


Figura 13-1. Un objeto moviéndose sobre una curva sin roce bajo la acción de la gravedad.

¿Funciona sólo en la caída libre o es más general? Esperamos de nuestra discusión sobre la conservación de la energía, que ésta dé resultados para un objeto que se mueve de un punto a otro en una especie de curva sin roce, bajo la influencia de la gravedad (figura 13-1). Si el objeto alcanza cierta altura h desde una altura original H, la misma fórmula sería correcta también, aunque la velocidad sea ahora en alguna dirección distinta de la vertical. Nos gustaría comprender por qué la ley es todavía correcta. Sigamos el mismo análisis, averiguando la derivada de la energía cinética respecto al tiempo. Será otra vez $mv(dv/dt)$, pero $m(dv/dt)$ es la derivada de la magnitud del momentum, esto es, la *fuerza en la dirección del movimiento*, la fuerza tangencial F_t. Así

$$\frac{dT}{dt} = mv\,\frac{dv}{dt} = F_t v.$$

Ahora bien, la velocidad es la derivada temporal de la distancia a lo largo de la curva, ds/dt, y la fuerza tangencial F_t no es $-mg$, sino que es menor en el cociente entre la distancia dh a lo largo del camino y la distancia vertical ds. En otras palabras,

$$F_t = -mg\operatorname{sen}\theta = -mg\,\frac{dh}{ds},$$

de manera que

$$F_t\,\frac{ds}{dt} = -mg\left(\frac{dh}{ds}\right)\left(\frac{ds}{dt}\right) = -mg\,\frac{dh}{dt},$$

ya que los ds se cancelan. Así, pues, tenemos $-mg(dh/dt)$, que es igual a la derivada de $-mgh$ como antes.

Para poder entender exactamente cómo funciona la conservación de la energía en la mecánica en general, discutiremos enseguida cierto número de conceptos que nos ayudarán a analizarlo.

Primero, discutiremos la variación de la energía cinética en general en tres dimensiones. La energía cinética en tres dimensiones es

$$T = \tfrac{1}{2}m(v_x^2 + v_y^2 + v_z^2).$$

Si derivamos esta ecuación respecto al tiempo, obtenemos tres términos espantosos:

$$\frac{dT}{dt} = m\left(v_x\,\frac{dv_x}{dt} + v_y\,\frac{dv_y}{dt} + v_z\,\frac{dv_z}{dt}\right). \tag{13.4}$$

Pero $m(dv_x/dt)$ es la fuerza F_x que actúa sobre el objeto en la dirección de x. De manera que el primer miembro de la ecuación (13.4) es $F_xv_x + F_yv_y + F_zv_z$. Recordando nuestro cálculo vectorial reconocemos esto como $\boldsymbol{F}\cdot\boldsymbol{v}$; por consiguiente,

$$dT/dt = \boldsymbol{F}\cdot\boldsymbol{v}. \tag{13.5}$$

Este resultado puede deducirse más rápidamente como sigue: si $\boldsymbol{a}$ y $\boldsymbol{b}$ son dos vectores que dependen del tiempo, la derivada de $\boldsymbol{a}\cdot\boldsymbol{b}$ es, en general,

$$d(\boldsymbol{a}\cdot\boldsymbol{b})/dt = \boldsymbol{a}\cdot(d\boldsymbol{b}/dt) + (d\boldsymbol{a}/dt)\cdot\boldsymbol{b}. \tag{13.6}$$

Entonces usamos esto en la forma $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{v}$:

$$\frac{d(\frac{1}{2}mv^2)}{dt} = \frac{d(\frac{1}{2}m\boldsymbol{v}\cdot\boldsymbol{v})}{dt} = m\,\frac{d\boldsymbol{v}}{dt}\cdot\boldsymbol{v} = \boldsymbol{F}\cdot\boldsymbol{v} = \boldsymbol{F}\cdot\frac{d\boldsymbol{s}}{dt}. \tag{13.7}$$

Debido a que los conceptos de energía cinética y energía en general son tan importantes, se han dado diversos nombres a los importantes términos en ecuaciones como éstas. $\frac{1}{2}mv^2$ se denomina, como sabemos, *energía cinética*. $\boldsymbol{F}\cdot\boldsymbol{v}$ se llama *potencia*: la fuerza que actúa sobre un objeto por la velocidad del objeto (producto escalar de vectores) es la potencia entregada al objeto por esa fuerza.

Así, pues, tenemos un maravilloso teorema: *la variación de la energía cinética de un objeto es igual a la potencia gastada por la fuerza que actúa sobre él.*

Sin embargo, para estudiar la conservación de la energía queremos analizar esto más detenidamente. Calculemos la variación de la energía cinética en un tiempo dt muy corto. Si multiplicamos ambos miembros de la ecuación (13.7) por dt, encontramos que la variación diferencial de la energía cinética es la fuerza "punto" la distancia diferencial en que se ha movido

$$dT = \boldsymbol{F} \cdot d\boldsymbol{s}. \tag{13.8}$$

Si ahora integramos, obtenemos,

$$\Delta T = \int_1^2 \boldsymbol{F} \cdot d\boldsymbol{s}. \tag{13.9}$$

¿Qué significa esto? Significa que si un objeto se mueve *de una manera cualquiera* bajo la influencia de una fuerza, moviéndose en alguna clase de trayectoria curva, entonces la variación de EC cuando va de un punto a otro a lo largo de la curva es igual a la integral de la componente de la fuerza según la curva, por el desplazamiento diferencial ds, efectuándose la integración de un punto al otro. Esta integral también tiene nombre; se llama *trabajo efectuado por la fuerza sobre el objeto.* Vemos inmediatamente que *potencia es trabajo realizado por segundo.* Vemos también que es sólo una componente de la fuerza *en la dirección del movimiento* la que contribuye al trabajo efectuado. En nuestro sencillo ejemplo, las fuerzas eran sólo verticales y tenían una sola componente, digamos F_z, igual a $-mg$. No importa cómo se mueve el objeto en esas circunstancias, cayendo en parábola por ejemplo, $\boldsymbol{F} \cdot \boldsymbol{s}$, que se puede escribir como: $F_x dx + F_y dy + F_z dz$, se reduce a $F_z dz = -mgdz$, porque las otras componentes de la fuerza son cero. Por consiguiente, en nuestro sencillo caso:

$$\int_1^2 \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{z_1}^{z_2} -mg\,dz = -mg(z_2 - z_1), \tag{13.10}$$

de manera que otra vez descubrimos que sólo la *altura vertical* desde la que cae el objeto es la que contribuye a la energía potencial.

Una palabra sobre unidades. Puesto que las fuerzas se miden en newtons, y multiplicamos por una distancia para obtener trabajo, el trabajo se mide en *newton · metro* (n · m), pero a la gente no le gusta decir newton-metros, sino que prefiere decir *joules* (J). Un newton-metro se denomina joule: el trabajo se mide en joules. Potencia, entonces, es joules por segundo y a esto se le llama *watt* (W). Si multiplicamos watts por el tiempo, el resultado es el trabajo efectuado. El trabajo que efectúa la compañía eléctrica en nuestras casas, técnicamente hablando, es igual a los watts por el tiempo. De ahí obtenemos cosas como kilowatt-hora, 1 000 watts por 3 600 segundos o 3.6×10^6 joules.

Tomemos ahora otro ejemplo de la ley de conservación de la energía. Consideramos un objeto que tiene inicialmente energía cinética y que se mueve muy rápido y que resbala sobre el piso con fricción. Se detiene. Al comienzo la energía cinética *no* es cero, pero al final *es* cero; hay trabajo efectuado por las fuerzas, porque cuando hay fricción siempre hay una componente de la fuerza en dirección opuesta a la del movimiento, de manera que la energía se pierde paulatinamente. Pero ahora consideremos una masa ubicada en el extremo de un pivote que oscila sin fricción en un plano vertical en un campo gravitacional. Lo que sucede aquí es diferente, porque cuando la masa se mueve hacia arriba, la fuerza actúa hacia abajo y cuando la masa baja, la fuerza también es hacia abajo. Así $\boldsymbol{F} \cdot d\boldsymbol{s}$ conserva un signo al moverse hacia arriba y otro al bajar. En cada punto correspondiente en el camino hacia arriba y hacia abajo, los valores de $\boldsymbol{F} \cdot d\boldsymbol{s}$ son exactamente iguales en magnitud, pero de signos opuestos, de manera que el resultado neto de la integral será cero para este caso. La energía cinética con que la masa regresa al punto más bajo es la misma que tenía al partir; esto es el principio de la conservación de la energía. (Noten que cuando existen fuerzas de fricción, parece a primera vista que la conservación de la energía no es válida.

Tenemos que encontrar otra *forma* de energía. Resulta, en realidad, que se genera *calor* en un objeto cuando frota con otro, pero por el momento, según lo que se supone, eso no lo sabemos.)

13-2 Trabajo realizado por la gravedad

El siguiente problema a discutir es mucho más difícil que el anterior; tiene que ver con el caso en que las fuerzas no son constantes o simplemente verticales, como lo fueron en los casos que acabamos de resolver. Queremos considerar un planeta, por ejemplo, que se mueve alrededor del Sol o un satélite en el espacio alrededor de la Tierra.

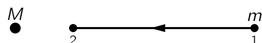


Figura 13-2. Una masa pequeña m cae bajo la acción de la gravedad hacia la masa grande M.

Primero consideremos el movimiento de un objeto que parte desde el punto 1 y cae, digamos, *directamente* hacia el Sol o hacia la Tierra (figura 13-2). ¿Habrá una ley de conservación de la energía en estas circunstancias? La única diferencia es que en este caso la fuerza está *cambiando* a medida que avanzamos, no es precisamente una constante. Como sabemos, la fuerza es $-GM/r^2$ multiplicado por la masa m, donde m es la masa que se mueve. Ahora, ciertamente, cuando un cuerpo cae a la Tierra, la energía cinética aumenta al incrementarse la distancia caída, como sucede cuando no nos preocupamos de la variación de la fuerza con la altura. El asunto es, si es posible encontrar otra fórmula para la energía potencial diferente de mgh, una función diferente de la distancia a la Tierra, para que la conservación de la energía continúe siendo válida. Este caso unidimensional es fácil de resolver porque sabemos que la variación de la energía cinética es igual a la integral, de un extremo del movimiento al otro, de $-GMm/r^2$ por el desplazamiento dr

$$T_2 - T_1 = -\int_1^2 GMm\,\frac{dr}{r^2}. \tag{13.11}$$

No se necesitan cosenos en este caso porque la fuerza y el desplazamiento tienen la misma dirección. Es fácil integrar dr/r^2; el resultado es $-1/r$, de tal manera que la ecuación (13.11) se transforma en

$$T_2 - T_1 = +GMm\left(\frac{1}{r_2} - \frac{1}{r_1}\right). \tag{13.12}$$

Así obtenemos una fórmula diferente para la energía potencial. La ecuación (13.12) nos dice que la cantidad $(1/2mv^2 - GMm/r)$ calculada en el punto 1, en el punto 2 o en cualquier otro lugar, tiene un valor constante.

Ya tenemos la fórmula de la energía potencial para el movimiento vertical en un campo gravitacional. Ahora se nos presenta un problema interesante. ¿Podemos conseguir *movimiento perpetuo* en un campo gravitacional? El campo gravitacional varía; en lugares diferentes tiene direcciones diferentes y tiene intensidades distintas. ¿Podríamos hacer algo así como esto, usando una trayectoria fija y sin fricción: empezar en algún punto y levantar un objeto hasta algún otro punto, luego moverlo en un arco hasta un tercer punto, luego bajarlo cierta distancia, después moverlo oblicuamente y sacarlo por otro camino, para que al llevarlo al punto de partida cierta cantidad de trabajo haya sido efectuada por la fuerza gravitacional y la energía cinética del objeto haya aumentado? ¿Podemos idear la curva de manera que el objeto vuelva moviéndose un poco más rápido que antes, de manera que dé vueltas y vueltas y nos dé movimiento perpetuo? Como el movimiento perpetuo es imposible, debemos hallar que también esto es imposible. Debemos descubrir la siguiente proposición: como no hay fricción, el objeto debería volver con una velocidad ni mayor ni menor —debería poder seguir dando vueltas y vueltas en cualquier trayectoria cerrada—. Expresándolo de otra manera, *el trabajo total efectuado al completar un ciclo debe ser cero* para fuerzas de gravedad, porque si no es cero, podemos obtener energía dando vueltas. (Si el trabajo resulta ser menor que cero, de manera que obtenemos menos velocidad al dar vuelta en una dirección, entonces simplemente damos vuelta en la otra dirección, porque las fuerzas, por supuesto, dependen sólo de

la posición, no de la dirección; si en un sentido es positivo, en el otro sentido será negativo, de manera que, a menos que sea cero, obtendremos movimiento perpetuo dando vueltas en cualquier sentido.)

¿Es el trabajo realmente cero? Tratemos de demostrar que lo es. Primero vamos a explicar más o menos por qué es cero, y después lo examinaremos un poco mejor matemáticamente. Supongamos que empleamos una trayectoria sencilla como la que se ve en la figura 13-3, en que una masa pequeña se lleva del punto 1 al 2 y después da vuelta en un círculo hasta 3, sigue hacia 4, después a 5, 6, 7 y 8 y finalmente, vuelva a 1. Todas las líneas o son puramente radiales o circulares con M como centro. ¿Cuánto trabajo se realiza al llevar m por este camino? Entre los puntos 1 y 2 es GMm por la diferencia de $1/r$ entre estos dos puntos:

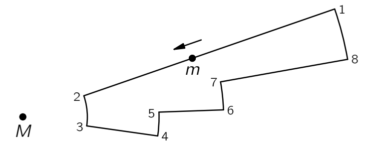


Figura 13-3. Una trayectoria cerrada en un campo gravitacional.

$$W_{12} = \int_1^2 \boldsymbol{F} \cdot d\boldsymbol{s} = \int_1^2 -GMm\,\frac{dr}{r^2} = GMm\left(\frac{1}{r_2} - \frac{1}{r_1}\right).$$

De 2 a 3 la fuerza es exactamente perpendicular a la curva, luego $W_{23} \equiv 0$. El trabajo entre 3 y 4 es

$$W_{34} = \int_3^4 \boldsymbol{F} \cdot d\boldsymbol{s} = GMm\left(\frac{1}{r_4} - \frac{1}{r_3}\right).$$

De la misma forma encontramos que $W_{45} = 0$, $W_{56} = GMm(1/r_6 - 1/r_5)$, $W_{67} = 0, W_{78} = GMm(1/r_8 - 1/r_7)$ y $W_{81} = 0$. Por consiguiente

$$W = GMm\left(\frac{1}{r_2} - \frac{1}{r_1} + \frac{1}{r_4} - \frac{1}{r_3} + \frac{1}{r_6} - \frac{1}{r_5} + \frac{1}{r_8} - \frac{1}{r_7}\right).$$

Pero notamos que $r_2 = r_3, r_4 = r_5, r_6 = r_7, r_8 = r_1$. Por consiguiente $W = 0$.

Naturalmente podríamos preguntarnos si ésta es una curva demasiado sencilla. ¿Qué pasa si consideramos una curva *real*? Hagamos la prueba con una curva real. Antes que nada podríamos afirmar que una curva real siempre se puede representar lo suficientemente bien por una serie de dientes de sierra como los que se ven en la figura 13-4 y que por consiguiente, etc., Q.E.D., pero sin un pequeño análisis no es obvio al principio, que el trabajo efectuado al recorrer aún un pequeño triángulo sea cero. Ampliemos uno de los triángulos, como se muestra en la figura 13-4. ¿Será el trabajo que se efectúa al moverse desde a hasta b y desde b hasta c en un triángulo el mismo que el trabajo efectuado al ir directamente desde a hasta c? Supongan que la fuerza actúa en cierta dirección; consideremos un triángulo cuyo lado bc tiene esta dirección, sólo como un ejemplo. También supongamos que el triángulo sea tan pequeño que la fuerza sea esencialmente constante en todo el triángulo. ¿Cuánto trabajo se efectúa al moverse desde a hasta c? Es:

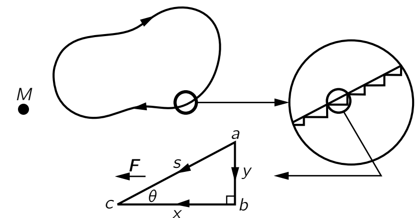


Figura 13-4. Una trayectoria cerrada "lisa" mostrando un segmento ampliado de ella aproximado con una serie de escalones formados por segmentos radiales y arcos de circunferencia, y una vista aumentada de un escalón.

$$W_{ac} = \int_a^c \boldsymbol{F} \cdot d\boldsymbol{s} = Fs\cos\theta,$$

ya que la fuerza es constante. Calculemos ahora el trabajo que se efectúa al recorrer los otros dos lados del triángulo. En el lado vertical ab la fuerza es perpendicular a ds, por lo que el trabajo aquí es cero. En el lado horizontal bc,

$$W_{bc} = \int_b^c \boldsymbol{F} \cdot d\boldsymbol{s} = Fx.$$

Entonces hemos visto que el trabajo efectuado al moverse a lo largo de los lados de un triángulo pequeño es igual al efectuado al moverse en un plano inclinado, porque $s\cos\theta$ es igual a x. Hemos demostrado previamente que el resultado es cero para cualquier trayectoria compuesta de una serie de muescas como las de la figura 13-3 y también que se efectúa el mismo trabajo si atravesamos de esquina a esquina en vez de seguir las muescas (siempre que las muescas sean lo suficientemente finas, y siempre podemos hacerlas muy finas); por consiguiente, *el trabajo efectuado al recorrer cualquier trayectoria cerrada en un campo gravitacional es cero.*

Éste es un resultado muy notable. Nos dice algo que no sabíamos antes respecto al movimiento planetario. Nos dice que cuando un planeta se mueve

alrededor del Sol (sin que haya otros cuerpos cerca, ni otras fuerzas), se mueve en tal forma, que el cuadrado de la velocidad en cualquier punto, menos ciertas constantes divididas entre el radio en ese punto, es siempre igual en cada punto de su órbita. Por ejemplo, cuanto más cerca del Sol se encuentre el planeta, más rápido se moverá; pero ¿cuánto más? Tanto como lo que sigue: si en vez de dejar que el planeta gire alrededor del Sol, le cambiáramos la dirección (pero no la magnitud) de su velocidad y lo hiciéramos mover radialmente y luego lo dejáramos caer desde un radio especial al radio que interese, la nueva velocidad sería la misma que la que tenía en su órbita real, porque éste es sólo otro ejemplo de trayectoria complicada. Siempre que regresemos a la misma distancia, la energía cinética será la misma. Así, tanto si el movimiento es el real, no perturbado, como si se le cambia la dirección por medio de canales, o movimientos de ligadura sin fricción, la energía cinética con que el planeta llega a cierto punto será siempre la misma.

Por lo tanto, cuando hacemos un análisis numérico del movimiento del planeta en su órbita, como lo hicimos anteriormente, podemos verificar en cada paso si estamos o no cometiendo errores apreciables al calcular esta cantidad constante, la energía, y ésta no debería cambiar. Para la órbita de la tabla 9-2 la energía sí varía,* en alrededor de 1.5 % desde el principio hasta el fin. ¿Por qué? Bien porque en el método numérico usamos intervalos no infinitesimales, o bien porque cometimos un pequeño error en los cálculos aritméticos en alguna parte.

Estudiemos la energía en otro caso: el problema de una masa fija a un resorte. Cuando desplazamos la masa de su posición de equilibrio, la fuerza restauradora es proporcional al desplazamiento. En esas circunstancias, ¿podemos deducir cierta ley para la conservación de la energía? Sí, porque el trabajo realizado por una fuerza como ésa

$$W = \int_0^x F\,dx = \int_0^x -kx\,dx = -\tfrac{1}{2}kx^2. \tag{13.13}$$

Por consiguiente, para una masa fija a un resorte tenemos que la energía cinética de la masa que oscila más $\frac{1}{2}kx^2$ es una constante. Veamos cómo funciona esto. Tiramos de la masa hacia abajo; está detenida aún, por lo que su velocidad es cero. Pero x no es cero, x está en su valor máximo, entonces existe alguna energía, la energía potencial por supuesto. Ahora soltamos la masa y algo empieza a suceder (no vamos a discutir detalles), pero en cualquier instante la energía cinética más la potencial debe ser una constante. Por ejemplo, una vez que la masa en su recorrido pasa el punto original de equilibrio, la posición x es igual a cero, pero esto ocurre cuando tiene su mayor v^2, y a medida que logra mayor x^2 obtiene menos v^2, y así sucesivamente. Por lo que el balance de x^2 y v^2 se mantiene cuando la masa se mueve hacia arriba y hacia abajo. Así tenemos ahora otra regla, que la energía potencial para un resorte es $\frac{1}{2}kx^2$, si la fuerza es $-kx$.

13-3 Suma de energías

Entramos ahora en consideraciones más generales sobre lo que sucede cuando hay un gran número de objetos. Supongan que tenemos el complicado problema constituido por muchos objetos, que designaremos por $i = 1, 2, 3, \ldots$, todos ejerciendo atracciones gravitacionales entre sí. ¿Qué sucede entonces? Vamos a demostrar que si sumamos las energías cinéticas de todas las partículas y agregamos a ésta la suma, extendida a todos los *pares* de partículas de su energía potencial gravitatoria mutua, $-GMm/r_{ij}$, el total es una constante:

$$\sum_i \tfrac{1}{2}m_i v_i^2 + \sum_{(\text{pares } ij)} -\frac{Gm_i m_j}{r_{ij}} = \text{constante}. \tag{13.14}$$

¿Cómo lo demostramos? Derivamos cada miembro respecto al tiempo y obtenemos cero. Al derivar $\frac{1}{2}m_i v_i^2$, encontramos derivadas de la velocidad que son las fuerzas como en la ecuación (13.5). Remplazamos estas fuerzas por la ley de fuerza que

* La energía por unidad de masa es $\frac{1}{2}(v_x^2 + v_y^2) - 1/r$ en las unidades de la tabla 9-2.

conocemos a partir de la ley de gravedad de Newton y observamos entonces que lo que queda es menos la derivada respecto al tiempo de

$$\sum_{\text{pares}} -\frac{Gm_im_j}{r_{ij}}.$$

La derivada respecto al tiempo de la energía cinética es

$$\begin{aligned}\frac{d}{dt}\sum_i \tfrac{1}{2}m_iv_i^2 &= \sum_i m_i\frac{d\boldsymbol{v}_i}{dt}\cdot\boldsymbol{v}_i\\ &= \sum_i \boldsymbol{F}_i\cdot\boldsymbol{v}_i\\ &= \sum_i\left(\sum_j -\frac{Gm_im_j\boldsymbol{r}_{ij}}{r_{ij}^3}\right)\cdot\boldsymbol{v}_i.\end{aligned}\tag{13.15}$$

La derivada respecto al tiempo de la energía potencial es

$$\frac{d}{dt}\sum_{\text{pares}} -\frac{Gm_im_j}{r_{ij}} = \sum_{\text{pares}}\left(+\frac{Gm_im_j}{r_{ij}^2}\right)\left(\frac{dr_{ij}}{dt}\right).$$

Pero

$$r_{ij} = \sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2},$$

de manera que

$$\begin{aligned}\frac{dr_{ij}}{dt} &= \frac{1}{2r_{ij}}\bigg[2(x_i-x_j)\left(\frac{dx_i}{dt}-\frac{dx_j}{dt}\right)\\ &\qquad + 2(y_i-y_j)\left(\frac{dy_i}{dt}-\frac{dy_j}{dt}\right)\\ &\qquad + 2(z_i-z_j)\left(\frac{dz_i}{dt}-\frac{dz_j}{dt}\right)\bigg]\\ &= \boldsymbol{r}_{ij}\cdot\frac{\boldsymbol{v}_i-\boldsymbol{v}_j}{r_{ij}}\\ &= \boldsymbol{r}_{ij}\cdot\frac{\boldsymbol{v}_i}{r_{ij}} + \boldsymbol{r}_{ji}\cdot\frac{\boldsymbol{v}_j}{r_{ji}},\end{aligned}$$

ya que $\boldsymbol{r}_{ij} = -\boldsymbol{r}_{ji}$ mientras que $r_{ij} = r_{ji}$. Así

$$\frac{d}{dt}\sum_{\text{pares}} -\frac{Gm_im_j}{r_{ij}} = \sum_{\text{pares}}\left[\frac{Gm_im_j\boldsymbol{r}_{ij}}{r_{ij}^3}\cdot\boldsymbol{v}_i + \frac{Gm_jm_i\boldsymbol{r}_{ji}}{r_{ji}^3}\cdot\boldsymbol{v}_j\right].\tag{13.16}$$

Ahora debemos tomar nota cuidadosamente de lo que significan $\sum_i\{\sum_j\}$ y $\sum_{\text{pares}}$. En la ecuación (13.15) $\sum_i\{\sum_j\}$ significa que i toma todos los valores $i = 1, 2, 3, \ldots$ por turno, y para cada valor de i, el índice j toma todos los valores excepto i. Así, si $i = 3$, j toma los valores 1, 2, 4, …

En la ecuación (13.16), por otro lado, $\sum_{\text{pares}}$ significa que valores dados de i y de j ocurren sólo una vez. Así, el par de partículas 1 y 3 contribuye sólo con un término a la suma. Para seguir el hilo de esto, podemos ponernos de acuerdo que i asuma todos los valores 1, 2, 3, …, y que para cada i dejemos que j se extienda sólo sobre valores *mayores* que i. Entonces, si $i = 3$, j podría tomar sólo los valores 4, 5, 6… Pero notamos que para cada valor de i, j hay dos contribuciones a la suma, una en que interviene $\boldsymbol{v}_i$ y la otra en que interviene $\boldsymbol{v}_j$, y que estos términos tienen la misma apariencia que los de la ecuación (13.15), en que *todos* los valores de i y de j (excepto $i = j$) están incluidos en la suma. Por

consiguiente, al aparear los términos uno por uno, observamos que las ecuaciones (13.16) y (13.15) son exactamente iguales, pero de signo contrario, de manera que la derivada respecto al tiempo de la energía cinética más la energía potencial es en realidad cero. Así observamos que, para muchos objetos, *la energía cinética es la suma de las contribuciones de cada objeto individual*, y que la energía potencial es también sencilla, siendo también sólo una suma de contribuciones, las energías entre todos los pares. Podemos comprender *por qué* debe ser así la energía de cada par: supongan que queremos encontrar la cantidad total de trabajo que se debe efectuar para traer los objetos a ciertas distancias entre sí. Podemos hacer esto en varias etapas, trayéndolos desde el infinito, donde no existen fuerzas, uno por uno. Primero traemos al número uno, que no requiere trabajo, ya que no hay otros cuerpos presentes que ejerzan fuerzas sobre él. A continuación traemos al número dos, que requiere algo de trabajo, a saber $W_{12} = -Gm_1m_2/r_{12}$. Ahora, y éste es un punto importante, supongan que traemos el objeto siguiente a la posición 3. En cualquier momento la fuerza que actúa sobre el número 3 puede escribirse como la suma de dos fuerzas —la fuerza ejercida por el número 1 y la ejercida por el número 2—. Por consiguiente, *el trabajo efectuado es la suma de los trabajos efectuados por cada uno*, porque si $\boldsymbol{F}_3$ se puede descomponer en la suma de dos fuerzas,

$$\boldsymbol{F}_3 = \boldsymbol{F}_{13} + \boldsymbol{F}_{23},$$

entonces el trabajo es

$$\int \boldsymbol{F}_3 \cdot d\boldsymbol{s} = \int \boldsymbol{F}_{13} \cdot d\boldsymbol{s} + \int \boldsymbol{F}_{23} \cdot d\boldsymbol{s} = W_{13} + W_{23}.$$

Esto es, el trabajo efectuado es la suma del trabajo efectuado en contra de la primera fuerza y la segunda fuerza, como si cada una actuara independientemente. Procediendo de esta manera, vemos que el trabajo total que se necesita para armar la configuración dada de los objetos es precisamente el valor dado en la ecuación (13.14) como energía potencial. Es porque la gravedad obedece al principio de superposición de las fuerzas que podemos escribir la energía potencial como una suma sobre cada par de partículas.

13-4 Campo gravitacional de objetos grandes

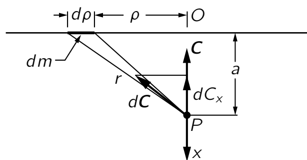


Figura 13-5. El campo gravitacional $\boldsymbol{C}$ sobre una masa puntual producida por una lámina plana infinita de materia.

Ahora vamos a calcular los campos que se encuentran en algunas circunstancias físicas en que interviene una *distribución de masa.* Hasta aquí no hemos considerado la distribución de masas, sólo partículas, de manera que es interesante calcular las fuerzas cuando son producidas por más de una sola partícula. Primero encontraremos la fuerza gravitacional sobre una masa producida por una hoja plana de material, de extensión infinita. La fuerza sobre una unidad de masa en un punto dado P, producida por esta hoja de material (figura 13-5), estará, por supuesto, dirigida hacia la hoja. Sea a la distancia del punto a la hoja y sea μ la masa por unidad de superficie de esta gran hoja. Supondremos que μ es constante, es una hoja de material uniforme. Ahora bien, ¿qué campo pequeño $d\boldsymbol{C}$ es producido por la masa dm que se encuentra entre ρ y $\rho + d\rho$, desde el punto O de la hoja más cercano al punto P? Respuesta: $d\boldsymbol{C} = -G(dm\boldsymbol{r}/r^3)$. Pero este campo está dirigido según $\boldsymbol{r}$, y sabemos que solamente quedará la componente x al sumar todos los pequeños vectores $d\boldsymbol{C}$ para producir $\boldsymbol{C}$. La componente x de $d\boldsymbol{C}$ es

$$dC_x = -G\,\frac{dm\,r_x}{r^3} = -G\,\frac{dm\,a}{r^3}.$$

Ahora bien, todas las masas dm que se encuentran a la misma distancia r de P darán la misma dC_x, de manera que podemos escribir inmediatamente que dm es la masa total del *anillo* entre ρ y $\rho + d\rho$, es decir, $dm = \mu 2\pi\rho d\rho$ ($2\pi\rho d\rho$ es el área de un anillo de radio ρ y ancho $d\rho$, si $d\rho \ll \rho$). Por lo tanto

$$dC_x = -G\mu 2\pi\rho\,\frac{d\rho\,a}{r^3}.$$

Entonces ya que $r^2 = \rho^2 + a^2, \rho d\rho = rdr$. Por consiguiente,

$$C_x = -2\pi G\mu a \int_a^\infty \frac{dr}{r^2} = -2\pi G\mu a\Big(\frac{1}{a} - \frac{1}{\infty}\Big) = -2\pi G\mu. \tag{13.17}$$

Luego, ¡la fuerza es independiente de la distancia a! ¿Por qué? ¿Hemos cometido algún error? Se podría pensar que mientras más nos alejamos, más débil sería la fuerza. ¡Pero no! Si estamos cerca, la mayor parte de la materia está atrayendo en un ángulo desfavorable; si estamos muy lejos, una mayor cantidad de materia está situada más favorablemente para ejercer una atracción hacia el plano. A cualquier distancia, la materia que es más efectiva se encuentra en cierto cono. Cuando estamos más lejos, la fuerza es menor según el cuadrado inverso, pero en el mismo cono, en el mismo ángulo, hay *mucha más materia*, ¡mayor precisamente en el cuadrado de la distancia! Este análisis puede hacerse riguroso sólo notando que la contribución diferencial en cualquier cono dado es en realidad independiente de la distancia, a causa de las variaciones inversas de la intensidad de la fuerza de una masa dada y la cantidad de masa incluida en el cono al cambiar la distancia. La fuerza no es realmente constante, por supuesto, porque al pasar al otro lado de la lámina cambia de signo.

En efecto, también hemos resuelto un problema eléctrico: si tenemos una placa cargada eléctricamente, con una carga σ por unidad de superficie, el campo eléctrico en un punto fuera de la placa es igual a $\sigma/2\epsilon_0$ y tiene dirección hacia afuera, si la placa está cargada positivamente, y hacia adentro, si la placa está cargada negativamente. Para demostrar esto, observemos sencillamente que $-G$, para la gravedad, juega el mismo papel que $\frac{1}{4}\pi\epsilon_0$ en la electricidad.

Supongan ahora que tenemos dos placas, con una carga positiva $+\sigma$ en una placa y una carga negativa $-\sigma$ en la otra a una distancia D de la primera. ¿Cuál es el campo? Fuera de las dos placas es cero. ¿Por qué? Porque una atrae y la otra repele, siendo la fuerza *independiente de la distancia*, ¡de manera que las dos se anulan! También, el campo *entre* las dos placas es claramente el doble de la que produce una placa, a saber $E = \sigma/\epsilon_0$, y está dirigida de la placa positiva a la negativa.

Ahora llegamos a un problema interesante e importante, cuya solución hemos estado suponiendo todo el tiempo, a saber, que la fuerza producida por la Tierra en un punto sobre su superficie o fuera de ella es la misma que se tendría si toda la masa de la Tierra estuviera situada en su centro. La validez de esta suposición no es evidente, porque cuando estamos cerca, parte de la masa se encuentra muy cerca de nosotros y parte se encuentra más lejos, y así sucesivamente. ¡Cuando sumamos todos los efectos, parece un milagro que la fuerza resultante sea igual a la que obtendríamos si colocáramos toda la masa en el centro!

Demostremos ahora que el milagro es correcto. Para hacer esto, sin embargo, vamos a considerar una cáscara hueca, delgada y uniforme en vez de la Tierra. Sea m la masa total de la cáscara, y calculemos la *energía potencial* de una partícula de masa m' situada a una distancia R de la esfera (figura 13-6) y demostremos que la energía potencial es la misma que si la masa m fuera un punto en el centro. (Es más fácil trabajar con la energía potencial que con el campo, porque no tenemos que preocuparnos de los ángulos, sólo sumamos las energías potenciales de todos los pedazos de masa.) Si designamos con x la distancia de cierta sección plana al centro, la masa que se encuentra en una rebanada dx está a la misma distancia r de P, y la energía potencial producida por este anillo es $-Gm'dm/r$. ¿Cuánta masa se encuentra en esta pequeña rebanada dx? Una cantidad

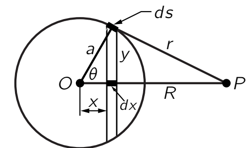


Figura 13-6. Una delgada capa esférica de masa o de carga.

$$dm = 2\pi y\mu\, ds = \frac{2\pi y\mu\, dx}{\operatorname{sen}\theta} = \frac{2\pi y\mu\, dx\, a}{y} = 2\pi a\mu\, dx,$$

donde $\mu = m/4\pi a^2$ es la densidad superficial de masa en la cáscara esférica, es una regla general que el área de una zona esférica es proporcional a su anchura axial. Por consiguiente, la energía potencial debida a dm es

$$dW = -\frac{Gm'\, dm}{r} = -\frac{Gm' 2\pi a\mu\, dx}{r}.$$

Pero vemos que

$$
\begin{aligned}
r^2 = y^2 + (R-x)^2 &= y^2 + x^2 + R^2 - 2Rx \\
&= a^2 + R^2 - 2Rx.
\end{aligned}
$$

Por lo tanto

$$2r\,dr = -2R\,dx$$

o

$$\frac{dx}{r} = -\frac{dr}{R}.$$

Por consiguiente,

$$dW = \frac{Gm'2\pi a\mu\,dr}{R}.$$

Entonces

$$
\begin{aligned}
W &= \frac{Gm'2\pi a\mu}{R}\int_{R+a}^{R-a} dr \\
&= -\frac{Gm'2\pi a\mu}{R}\,2a = -\frac{Gm'(4\pi a^2\mu)}{R} \\
&= -\frac{Gm'm}{R}.
\end{aligned}
\tag{13.18}
$$

Así, para una cáscaras esférica la energía potencial de una masa m' fuera de la esfera es la misma que se tendría si toda la masa estuviera concentrada en el centro. Podemos imaginar que la Tierra se compone de una serie de cáscaras esféricas concéntricas, cada una de las cuales contribuye a la energía, que depende sólo de su masa su la distancia al centro; sumándolas todas obtenemos la *masa total*, y por consiguiente, ¡la Tierra se comporta como si toda su materia estuviera concentrada en el centro!

Pero observen lo que ocurre si el punto se encuentra *dentro* de la esfera. Haciendo los mismos cálculos, pero con P en el interior, todavía obtenemos la diferencia de las dos r, pero ahora en la forma $a-R-(a+R) = -2R$, o sea el doble de la distancia al centro. En otras palabras, W se transforma en $W = -Gm'm/a$, que es *independiente* de R e independiente de la posición, es decir, encontramos la misma energía independientemente de *donde* nos encontremos en el interior. Por consiguiente, no hay fuerza; no se efectúa trabajo cuando nos movemos en el interior. Si la energía potencial es la misma, no importa dónde se coloca un objeto dentro de una esfera, no puede actuar ninguna fuerza sobre el objeto. Así es que no existe fuerza en el interior, sólo existe fuerza en el exterior y ésta es la misma que habría si toda la masa estuviera concentrada en el centro.

14

Trabajo y energía potencial (conclusión)

14-1 Trabajo

En el capítulo anterior hemos presentado muchas ideas y resultados que juegan un papel importante en la física. Estas ideas son tan importantes que parece que vale la pena dedicar un capítulo entero a examinarlas más detenidamente. En este capítulo no vamos a repetir las "demostraciones" o los trucos específicos por los que se obtuvieron los resultados, sino que más bien nos concentraremos en la discusión de las ideas.

En el aprendizaje de cualquier tema de naturaleza técnica en que la matemática juega un papel, uno se enfrenta a la tarea de entender y de almacenar en la memoria un enorme cuerpo de fenómenos e ideas, unidas por ciertas relaciones que puede "probarse" o "demostrarse" que existen entre ellas. Es fácil confundir la demostración misma con la relación que establece. Claramente, lo importante que hay que aprender y recordar es la relación, no la demostración. En cualquier circunstancia particular podemos, o decir "se puede demostrar que" esto o aquello es verdadero, o lo podemos demostrar. En casi todos los casos, la demostración en particular que se usa es elaborada, ante todo, de tal manera que se pueda escribir fácilmente en el pizarrón o papel, y en tal forma que aparezca lo más simple posible. En consecuencia, la demostración parece ser engañosamente sencilla cuando, en realidad, el autor puede haber trabajado durante horas, tratando de calcular lo mismo de diferentes maneras ¡hasta encontrar la manera más clara, para poder demostrarlo en el tiempo más corto posible! Lo que hay que recordar, al ver una demostración, no es la demostración misma, sino más bien que *se puede demostrar* que esto o aquello es verdadero. Por supuesto que cuando la demostración implica algunos procedimientos matemáticos, o "trucos", que uno no ha visto antes, se debe poner atención, no exactamente al "truco", sino a la idea matemática que envuelve.

Es seguro que en todas las demostraciones que se han hecho en un curso como éste, ninguna ha sido recordada desde el tiempo en que el autor estudió física de primer año. Todo lo contrario: él sólo recuerda que tal o cual es válida, y para explicar cómo algo puede ser demostrado inventa una demostración en el momento que la necesita. Cualquier persona que realmente ha aprendido una asignatura debería poder seguir un procedimiento similar, pero es inútil acordarse de las demostraciones. Por esta razón, en este capítulo evitaremos dar pruebas de las diversas afirmaciones hechas previamente, y sólo resumiremos los resultados.

La primera idea que se debe asimilar es *trabajo efectuado por una fuerza.* El término físico "trabajo" no es la palabra en el sentido ordinario como "¡trabajadores del mundo, únanse!", sino que es una idea diferente. Trabajo físico se expresa como $\int \boldsymbol{F} \cdot d\boldsymbol{s}$ y se llama "la integral de línea de $\boldsymbol{F}$ producto escalar $d\boldsymbol{s}$", que significa que si la fuerza, por ejemplo, tiene cierta dirección y el objeto sobre el cual se aplica se desplaza en otra dirección, entonces *sólo la componente de la fuerza en la dirección del desplazamiento* efectúa algún trabajo. Si, por ejemplo, la fuerza fuera constante y el desplazamiento fuera una distancia finita $\Delta\boldsymbol{s}$, entonces el trabajo efectuado al mover el objeto esa distancia, es sólo la componente de la fuerza en la dirección de $\Delta\boldsymbol{s}$ por Δs. La regla es "fuerza por distancia", pero en realidad queremos decir que sólo la componente de la fuerza en la dirección del desplazamiento por Δs, o equivalentemente, la componente del desplazamiento en la dirección de la fuerza por F. Es evidente que no se efectúa trabajo alguno por una fuerza que actúa en ángulo recto al desplazamiento.

Ahora, si el vector desplazamiento $\Delta \boldsymbol{s}$ se resuelve en componentes, en otras palabras, si el desplazamiento real es $\Delta \boldsymbol{s}$ y deseamos considerarlo efectivamente como una componente Δx del desplazamiento en la dirección x, Δy en la dirección y y Δz en la dirección z, entonces el trabajo efectuado al llevar un objeto de un lugar a otro se puede calcular en tres partes calculando el trabajo efectuado según x, según y y según z. El trabajo efectuado al moverse a lo largo de x involucra sólo esa componente de fuerza, es decir F_x y así sucesivamente, de manera que el trabajo es $F_x\Delta x + F_y\Delta y + F_z\Delta z$. Cuando la fuerza no es constante y tenemos un movimiento curvilíneo complicado, debemos resolver la trayectoria en un gran número de pequeñas $\Delta \boldsymbol{s}$, sumar el trabajo efectuado al mover el objeto en cada $\Delta \boldsymbol{s}$, y tomar el límite cuando $\Delta \boldsymbol{s}$ tiende a cero. Éste es el significado de "integral de línea".

Todo lo que acabamos de decir está contenido en la fórmula $W = \int \boldsymbol{F} \cdot d\boldsymbol{s}$. Está muy bien poder decir que es una fórmula maravillosa, pero otra cosa es comprender su significado, o cuáles son algunas de sus consecuencias.

La palabra "trabajo" en física tiene un significado tan diferente del de la palabra que se usa en circunstancias ordinarias, que debe observarse cuidadosamente que hay circunstancias peculiares en las que no es lo mismo. Por ejemplo, según la definición física de trabajo, si alguien sujeta por un rato un peso de 100 kilos a cierta altura del suelo, no está haciendo trabajo. Sin embargo, todos saben que empieza a transpirar, temblar y perder el aliento, como si subiera corriendo por una escalera. Sin embargo, correr escaleras arriba *se* considera trabajar (al correr *escalera abajo* uno obtiene trabajo del mundo, según la física), pero al sujetar simplemente un objeto en una posición fija, no se realiza trabajo. Claramente, la definición física de trabajo difiere de la definición fisiológica, por razones que exploraremos brevemente.

Es un hecho que cuando uno sujeta un peso, hace trabajo "fisiológico". ¿Por qué habría de sudar? ¿Por qué se necesita consumir alimentos para sujetar el peso? ¿Por qué está el mecanismo que tiene dentro de sí operando a toda marcha únicamente para sujetar el peso? En realidad, el peso podría sujetarse sin esfuerzo sólo colocándolo sobre una mesa, ¡entonces la mesa, calmadamente y sin ruidos, ni provisión de energía, puede mantener el mismo peso a la misma altura! La situación fisiológica es algo como lo que sigue. Hay dos clases de músculos en el cuerpo humano y en el de otros animales: una clase llamada músculo *estriado* o *esquelético*, es el tipo de músculo que tenemos en los brazos, por ejemplo, que están bajo control voluntario; la otra clase, llamamos músculos *lisos*, es como el músculo en los intestinos, o en la almeja el gran músculo aductor que cierra la concha. Los músculos lisos funcionan muy lentamente, pero pueden sujetar una "posición", es decir, si la almeja trata de cerrar su concha en cierta posición, mantendrá esa posición, aun cuando haya una fuerza grande que procura cambiarla. Mantendrá su posición bajo carga por horas y horas sin cansarse, porque es como una mesa que sujeta un peso; "fija" cierta posición y las moléculas sencillamente se traban por cierto tiempo sin efectuar trabajo, sin que la almeja genere esfuerzo alguno. El hecho de que tengamos que generar esfuerzo para sujetar un peso sencillamente se debe al diseño del músculo estriado. Lo que sucede es que cuando un impulso nervioso llega hasta una fibra muscular, la fibra da una pequeña contracción y después se relaja, de manera que cuando sujetamos algo, descargas enormes de impulsos nerviosos llegan al músculo, un gran número de contracciones sujetan el peso, mientras que las otras fibras se relajan. Podemos ver esto, por supuesto: cuando sujetamos un objeto pesado y nos cansamos, empezamos a temblar. La razón es que las descargas llegan irregularmente, y el músculo está cansado y no reacciona lo suficientemente rápido. ¿Por qué tenemos un diseño tan ineficiente? No sabemos bien por qué, pero la evolución no ha podido desarrollar músculos lisos *rápidos*. El músculo liso sería mucho más efectivo para sostener pesos, porque podría pararse ahí y se fijaría; no involucraría trabajo y no se necesitaría energía. Sin embargo, tiene el inconveniente de que funciona demasiado lento.

Volviendo ahora a la física, podemos preguntar *por qué* queremos calcular el trabajo efectuado. La respuesta es que es interesante y útil hacerlo, ya que

el trabajo efectuado sobre una partícula por la resultante de todas las fuerzas que actúan sobre ella es exactamente igual al cambio de energía cinética de la partícula. Esto es, si se empuja un objeto, aumenta su velocidad y

$$\Delta(v^2) = \frac{2}{m}\,\boldsymbol{F}\cdot\Delta\boldsymbol{s}.$$

14-2 Movimientos con ligadura

Otro aspecto interesante de las fuerzas y el trabajo es éste: supóngase que tenemos una pista inclinada o curvilínea y tenemos una partícula que debe moverse en esta pista, pero sin fricción. O bien, podemos tener un péndulo formado por una cuerda y un peso; la cuerda obliga al peso a moverse en un círculo alrededor del punto de pivote. Se puede cambiar el punto de pivote haciendo que la cuerda golpee un obstáculo de manera que la trayectoria del peso sigue en dos círculos de diferentes radios. Éstos son ejemplos de lo que llamamos *ligaduras fijas sin fricción.*

En movimientos con una ligadura fija sin fricción no se hace trabajo por la ligadura, porque las fuerzas de ligadura son siempre perpendiculares al movimiento. Por "fuerzas de ligadura" entendemos fuerzas que la propia ligadura aplica directamente: la fuerza de contacto con la pista, o la tensión en la cuerda.

Las fuerzas que intervienen en el movimiento de una partícula que se mueve sobre una pendiente bajo la influencia de la gravedad son bastante complicadas, porque hay una fuerza vincular, una fuerza gravitacional, etc. Sin embargo, si basamos nuestros cálculos del movimiento en la conservación de la energía y *la fuerza gravitacional sola*, obtenemos el resultado correcto. Esto parece bastante extraño, porque no es estrictamente la manera correcta de hacerlo —debiéramos usar la fuerza *resultante*—. No obstante, el trabajo efectuado por la fuerza gravitacional sola resultaría ser el cambio de la energía cinética, porque el trabajo efectuado por la parte que corresponde a la fuerza de vínculo es cero (figura 14-1).

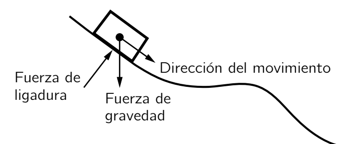


Figura 14-1. Fuerzas que actúan sobre un cuerpo que se desliza (sin fricción).

El aspecto importante en esto es que si una fuerza puede ser analizada como la suma de dos o más "partes", el trabajo efectuado por la fuerza resultante, al moverse en cierta trayectoria curvilínea, es la suma de los trabajos efectuados por las diversas "componentes" de las fuerzas en que la fuerza ha sido descompuesta. Así, si analizamos la fuerza como la suma vectorial de varios efectos, gravitacional más fuerzas de ligadura, etc., o la componente x de todas las fuerzas y la componente y de todas las fuerzas, o de cualquier otra manera que queramos dividirla, entonces el trabajo efectuado por la fuerza resultante es igual a la suma de los trabajos efectuados por todas las partes en que hemos dividido la fuerza al hacer la descomposición.

14-3 Fuerzas conservativas

En la naturaleza hay ciertas fuerzas, la de gravedad por ejemplo, que tiene una característica muy notable que llamamos "conservativa" (nada tiene que ver con ideas políticas, es otra vez una de esas "palabras locas"). Si calculamos cuánto trabajo efectúa una fuerza al mover un objeto de un punto a otro siguiendo una trayectoria curva, en general el trabajo depende de la curva, pero en casos especiales no. Si no depende de la curva, decimos que la fuerza es una fuerza conservativa. En otras palabras, si la integral de la fuerza multiplicada por la distancia recorrida entre la posición 1 a la posición 2 en la figura 14-2 se calcula a lo largo de la curva A y después a lo largo de B, obtenemos el mismo número de joules, y si esto es verdad para este par de puntos en *toda curva* y si la misma proposición se cumple, *cualquiera que sea el par de puntos* que usemos, entonces decimos que la fuerza es conservativa. En tales circunstancias, la integral del trabajo al ir de 1 a 2 puede ser calculada de una manera sencilla y podemos dar una fórmula para el resultado. En realidad no es tan fácil, porque también tenemos que especificar la curva, pero cuando tenemos un caso en que el trabajo

no depende de la curva, entonces, por supuesto, el trabajo depende sólo de las *posiciones* 1 y 2.

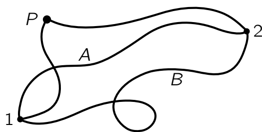


Figura 14-2. Trayectorias posibles entre dos puntos en un campo de fuerza.

Para demostrar esta idea, consideremos lo siguiente. Consideramos un punto P de referencia, en una ubicación arbitraria (figura 14-2). Entonces, la integral de línea de trabajo de 1 a 2, que deseamos calcular, puede ser evaluada como el trabajo efectuado al ir de 1 a P más el trabajo efectuado en ir de P a 2, porque las fuerzas son conservativas y el trabajo no depende de la curva. Ahora bien, el trabajo efectuado al ir de la posición P a una posición particular en el espacio es una función de esa posición en el espacio. Por supuesto, depende también de P, pero mantenemos el punto arbitrario P fijo permanentemente para el análisis. Si se hace eso, el trabajo efectuado al moverse el punto P al punto 2 es cierta función de la posición final de 2. Depende de dónde está 2; si vamos a algún otro punto, obtendremos un resultado diferente.

Llamaremos esta función de posición $U(x, y, z)$, y cuando queramos referirnos a algún punto en particular 2 cuyas coordenadas son (x_2, y_2, z_2), escribiremos $U(2)$, como abreviación de $U(x_2, y_2, z_2)$. El trabajo efectuado al ir del punto 1 al punto P también puede escribirse yendo *en el otro sentido* a lo largo de la integral, invirtiendo todas las $d\boldsymbol{s}$. Esto es, el trabajo efectuado al ir de 1 a P es *menos* el trabajo efectuado al ir del punto P al punto 1:

$$\int_1^P \boldsymbol{F} \cdot d\boldsymbol{s} = \int_P^1 \boldsymbol{F} \cdot (-d\boldsymbol{s}) = -\int_P^1 \boldsymbol{F} \cdot d\boldsymbol{s}.$$

Así, el trabajo efectuado al ir de P a 1 es $-U(1)$, y de P a 2 el trabajo es $-U(2)$. Por consiguiente, la integral de 1 a 2 es igual a $-U(2)$ más $[-U(1)$ invertido$]$ o $+U(1) - U(2)$:

$$U(1) = -\int_P^1 \boldsymbol{F} \cdot d\boldsymbol{s}, \qquad U(2) = -\int_P^2 \boldsymbol{F} \cdot d\boldsymbol{s},$$

$$\int_1^2 \boldsymbol{F} \cdot d\boldsymbol{s} = U(1) - U(2). \tag{14.1}$$

La cantidad $U(1) - U(2)$ se llama variación de energía potencial y llamamos energía potencial a U. Diremos que cuando el objeto está ubicado en la posición 2, tiene la energía potencial $U(2)$ y en la posición 1, tiene la energía potencial $U(1)$. Si el objeto se encuentra en P, tiene energía potencial cero. Si hubiéramos considerado cualquier otro punto, digamos Q en vez de P, resultaría (y dejaremos que ustedes lo demuestren), que *la energía potencial cambia sólo en el agregado de una constante*. Ya que la conservación de la energía depende sólo de las variaciones, no importa que le agreguemos una constante a la energía potencial. De manera que el punto P es arbitrario.

Tenemos ahora las dos siguientes proposiciones: (1) que el trabajo efectuado por una fuerza es igual a la variación de energía cinética de la partícula, pero (2) matemáticamente, para una fuerza conservativa, el trabajo efectuado es menos la variación de una función U que llamamos energía potencial. Como consecuencia de estas dos, llegamos a la proposición de que si *sólo actúan fuerzas conservativas, la energía cinética T más la energía potencial U permanece constante*:

$$T + U = \text{constante}. \tag{14.2}$$

Discutamos ahora las fórmulas para la energía potencial en varios casos. Si tenemos un campo gravitacional uniforme, si no ascendemos a alturas comparables al radio de la Tierra, la fuerza es una fuerza vertical constante y el trabajo efectuado es sencillamente la fuerza por la distancia vertical. Así,

$$U(z) = mgz, \tag{14.3}$$

y el punto P que corresponde a energía potencial cero resulta ser cualquier punto en el plano $z = 0$. También podríamos haber dicho que la energía potencial es

$mg(z-6)$ si lo hubiéramos deseado; todos los resultados serían, por supuesto, los mismos en nuestro análisis excepto que el valor de la energía potencial en $z=0$ sería $-6mg$. No importa, porque sólo las *diferencias* en energía potencial cuentan.

La energía que se necesita para comprimir un resorte lineal una distancia x del punto en equilibrio es

$$U(x)=\tfrac{1}{2}kx^2, \tag{14.4}$$

y el cero de energía potencial está en el punto $x=0$, la posición de equilibrio del resorte. Otra vez podríamos agregar cualquier constante que deseáramos.

La energía potencial de gravitación para puntos de masas M y m, separados por una distancia r, es

$$U(r)=-GMm/r. \tag{14.5}$$

Se ha escogido aquí la constante, tal que el potencial sea cero en el infinito. Además, la misma fórmula se aplica a cargas eléctricas, porque es la misma ley:

$$U(r)=q_1q_2/4\pi\epsilon_0 r. \tag{14.6}$$

Usemos ahora una de estas fórmulas en casos reales para ver si entendemos lo que significan. *Pregunta*: ¿Con qué velocidad debemos disparar un cohete desde la Tierra para que pueda escapar? *Solución*: la energía cinética más la potencial deben ser una constante; cuando "escapa" estará a una distancia de millones de kilómetros y si apenas puede escapar, podemos suponer que se mueve con velocidad cero allá lejos, apenas moviéndose. Sea a el radio de la Tierra y M su masa. La energía cinética más la potencial es entonces al principio $\frac{1}{2}mv^2-GmM/a$. Al final del movimiento las dos energías deben ser iguales. La energía cinética se considera cero al final del movimiento, porque se supone que apenas se aleja a una velocidad esencialmente cero, y la energía potencial es GmM dividido entre infinito, que es cero. De manera que todo es cero en un lado y eso nos dice que el cuadrado de la velocidad debe ser $2GM/a$. Pero GM/a^2 es lo que llamamos aceleración de gravedad g. Entonces

$$v^2=2ga.$$

¿A qué velocidad debe moverse un satélite para mantenerse girando alrededor de la Tierra? Hace mucho tiempo que resolvimos esto y encontramos que $v^2=GM/a$. Por consiguiente, para *alejarse* de la Tierra, necesitamos $\sqrt{2}$ veces la velocidad que necesitaríamos para girar *alrededor* de la Tierra cerca de su superficie. Se necesita, en otras palabras, *el doble de la energía* (porque la energía aumenta con el cuadrado de la velocidad) para escapar de la Tierra que para girar alrededor de ella. Por consiguiente, lo primero que históricamente se hizo con los satélites fue conseguir que uno se moviera alrededor de la Tierra, lo que requiere una velocidad de cinco millas por segundo. Lo siguiente fue enviar un satélite fuera de la Tierra permanentemente; esto requería el doble de la energía o cerca de siete millas por segundo.

Continuando ahora con nuestra discusión de las características de la energía potencial, consideremos la interacción de dos moléculas, o de dos átomos, dos átomos de oxígeno por ejemplo. Cuando están muy separados, la fuerza es de atracción, que varía inversamente con la séptima potencia de la distancia, y cuando están muy cerca, la fuerza es una repulsión muy grande. Si se integra la séptima potencia para encontrar el trabajo efectuado, encontramos que la energía potencial U, que es función de la distancia radial entre los dos átomos de oxígeno, varía inversamente con la sexta potencia de la distancia para distancias grandes.

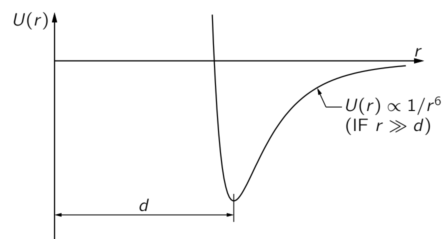


Figura 14-3. La energía potencial entre dos átomos en función de su distancia.

Si dibujamos la curva de la energía potencial $U(r)$, como en la figura 14-3, partimos así para r grande con la inversa de una sexta potencia, pero si nos acercamos lo suficiente, alcanzamos un punto d donde hay un mínimo de energía potencial. El mínimo de energía potencial en $r=d$ significa esto: si comenzamos en d y nos movemos una distancia corta, una distancia muy corta, el trabajo

efectuado, que es la variación de energía potencial cuando nos movemos esta distancia, es casi cero, porque hay muy poca variación de energía potencial en el fondo de la curva. Por tanto, no hay fuerza en este punto, de manera que es el punto de equilibrio. Otra manera de comprender que éste es el punto de equilibrio es que se necesita hacer trabajo para alejarse de d en cualquiera de las dos direcciones. Cuando los dos átomos de oxígeno se han ubicado, de manera que no se puede liberar más energía a partir de la fuerza entre ellos, están en el estado más bajo de energía y se encontrarán con esta separación d. Éste es la apariencia de una molécula de oxígeno cuando está fría. Cuando la calentamos, los átomos se agitan y se alejan y en realidad los podemos separar, pero para hacerlo se necesita cierta cantidad de trabajo o energía, que es la energía potencial entre $r = d$ y $r = \infty$. Cuando intentamos comprimir los átomos para juntarlos, la energía sube rápidamente, porque se repelen.

La razón por la cual mencionamos esto es que el concepto de *fuerza* no es particularmente apropiado en la mecánica cuántica; allí el concepto de *energía* es mucho más natural. Encontramos que, a pesar de que fuerzas y velocidades se "disuelven" y desaparecen al considerar las fuerzas más avanzadas entre la materia nuclear y entre las moléculas, etc., el concepto de energía permanece. Por consiguiente, encontramos curvas de energía potencial en los libros de mecánica cuántica, pero muy rara vez vemos una curva para la fuerza entre dos moléculas, porque en esas ocasiones la gente que hace análisis está pensando en términos de energía en vez de fuerza.

A continuación notemos que si varias fuerzas conservativas actúan sobre un objeto al mismo tiempo, la energía potencial del objeto es la suma de las energías potenciales de cada una de las fuerzas separadas. Ésta es la misma proposición que mencionamos antes, porque si la fuerza se puede representar como una suma vectorial de fuerzas, el trabajo efectuado por la fuerza total es la suma de los trabajos hechos por las fuerzas parciales, y se puede entonces analizar como variaciones de las energías potenciales de cada una separadamente. Así, pues, la energía potencial total es la suma de todas las pequeñas partes.

Podríamos generalizar esto al caso de un sistema de muchos objetos que interactúan, como Júpiter, Saturno, Urano, etc., u oxígeno, nitrógeno, carbono, etc., que interactúan en pares, debido a fuerzas, todas las cuales son conservativas. En estas circunstancias, la energía cinética en todo el sistema es sencillamente la suma de las energías cinéticas de todos los átomos particulares o planetas o lo que sea y la energía potencial del sistema es la suma, sobre los pares de partículas, de la energía potencial de la acción recíproca de un solo par, como si los otros no estuvieran presentes. (En realidad esto no sucede para las fuerzas moleculares y la fórmula es algo más complicada; es por cierto válida para la gravitación newtoniana y es válida como aproximación para las fuerzas moleculares. Para las fuerzas moleculares hay una energía potencial, pero a veces es una función más complicada de la posición de los átomos que simplemente una suma de términos de pares.) En el caso especial de la gravedad, por consiguiente, la energía potencial es la suma sobre todos los pares i y j de $-Gm_im_j/r_{ij}$, como se indicó en la ecuación (13.14). Ésta ecuación expresaba matemáticamente la siguiente proposición: que la energía cinética total más la energía potencial total no varía en el tiempo. A medida que los diversos planetas dan vueltas, giran, oscilan, etc., si calculamos la energía cinética total y la energía potencial total encontramos que la suma total permanece constante.

14-4 Fuerzas no conservativas

Hemos dedicado bastante tiempo al estudio de las fuerzas conservativas; ¿qué hay de las fuerzas no conservativas? Daremos una mirada más profunda que lo común a esto, y estableceremos que ¡no hay fuerzas no conservativas! En realidad, todas las fuerzas fundamentales en la naturaleza son conservativas. Esto no es una consecuencia de las leyes de Newton. En realidad, hasta donde Newton mismo sabía, las fuerzas podían ser no conservativas, como aparentemente lo es la fricción. Cuando decimos *aparentemente* lo es la fricción, estamos tomando

un punto de vista moderno, en que se ha descubierto que todas las fuerzas elementales, las fuerzas entre las partículas en el nivel más fundamental, son conservativas.

Por ejemplo, si analizamos un sistema como el gran cúmulo globular de estrellas cuya fotografía hemos visto, con sus miles de estrellas, todas interactuando, la fórmula para la energía potencial total es sencillamente un término, más otro término, etc., sumando sobre todos los pares de estrellas y la energía cinética es la suma de las energías cinéticas de todas y cada una de las estrellas. Pero el cúmulo globular, como un todo, también se mueve en el espacio, y si estuviéramos lo suficientemente lejos para no ver los detalles, se podría considerar como un solo objeto. Entonces, si se le aplican fuerzas, algunas de esas fuerzas podrían resultar en empujarlo hacia adelante como un todo y veríamos moverse el centro de todo el conjunto. Por otro lado, algunas de las fuerzas pueden ser, por así decirlo, "desperdiciadas" en aumentar la energía cinética o potencial de las "partículas" en el interior. Supongamos, por ejemplo, que la acción de estas fuerzas expande todo el cúmulo y hace que las partículas se muevan con mayor rapidez. La energía total del todo en realidad se conserva, pero visto de afuera con nuestros toscos ojos que no pueden percibir la confusión de los movimientos en el interior, y sólo pensando en la energía cinética del movimiento de todo el objeto como si fuera una sola partícula, aparecería como si la energía no se conservara, pero esto se debe a una falta de comprensión de lo que vemos. Y éste resulta ser el caso: la energía total del universo, cinética más potencial, es una constante si la examinamos con suficiente cuidado.

Cuando estudiamos la materia, en los más finos detalles al nivel atómico, no siempre es *fácil* separar la energía total de una cosa en dos partes, energía cinética y energía potencial, y esa separación no siempre es necesaria. *Casi* siempre es posible hacerlo, de manera que digamos que siempre *es* posible y que la suma de la energía potencial más cinética del mundo es constante. Por lo tanto, el total de energía potencial más cinética dentro del mundo entero es constante, y si el "mundo" es un pedazo de material aislado, la energía es constante si no existen fuerzas externas. Pero, como hemos visto, algo de la energía cinética y potencial de un objeto puede ser interna, por ejemplo, el movimiento molecular interno, en el sentido de que no lo notamos. Sabemos que en un vaso de agua todo se agita, todas las partes se mueven todo el tiempo, de manera que existe una cierta energía cinética interior, a la que generalmente no prestamos atención. No notamos el movimiento de los átomos que produce calor y entonces no la denominamos energía cinética, pero el calor es primordialmente energía cinética. La energía potencial interna puede también presentarse, por ejemplo, en la forma de energía química: cuando quemamos gasolina, se libera energía porque las energías potenciales de los átomos en el nuevo ordenamiento atómico están en un nivel más bajo que en el ordenamiento antiguo. No es posible tratar estrictamente el calor como energía cinética pura, porque se introduce un poco de la potencial, y viceversa para la energía química, de manera que las consideramos en conjunto y decimos que el total de la energía cinética y potencial en el interior del objeto es en parte calor, en parte energía química, etc. De todas maneras, todas estas formas diferentes de energía interna se consideran a veces como energía "perdida" en el sentido expresado más arriba; esto se aclarará más cuando estudiemos termodinámica.

Como otro ejemplo, cuando hay fricción no es verdad que la energía cinética se pierde, aunque un objeto que resbala se detiene y la energía cinética parece perderse. La energía cinética no se pierde porque, por supuesto, los átomos en el interior se agitan con una cantidad mayor de energía cinética que antes y aunque no podemos ver eso, lo podemos medir determinando la temperatura. Por supuesto, que si no tomamos en cuenta la energía térmica, entonces el teorema sobre la conservación de la energía aparecerá como falso.

Otra situación en que la conservación de la energía parece ser falsa es cuando estudiamos sólo una parte de un sistema. Naturalmente, el teorema de conservación de la energía parecerá no ser válido si algo interactúa con otra cosa en el exterior y omitimos el tomar en cuenta esa interacción.

En la física clásica, la energía potencial comprendía sólo la gravitación y la electricidad, pero ahora tenemos energía nuclear y otras energías también. La luz, por ejemplo, implicaría una nueva forma de energía en la teoría clásica, pero también podemos, si queremos, imaginar que la energía de la luz es la energía cinética de un fotón y entonces nuestra fórmula (14.2) todavía estaría bien.

14-5 Potenciales y campos

Discutiremos ahora algunos de los conceptos asociados con la energía potencial y con el concepto de *campo*. Supongamos tener dos grandes objetos A y B y un tercero muy pequeño que es gravitacionalmente atraído por los dos con cierta fuerza resultante $\boldsymbol{F}$. Ya hemos notado en el capítulo 12 que la fuerza gravitacional sobre una partícula puede escribirse como su masa, m, por otro vector $\boldsymbol{C}$, que depende sólo de la *posición* de la partícula:

$$\boldsymbol{F} = m\boldsymbol{C}.$$

Podemos analizar la gravitación, entonces, imaginando que existe cierto vector $\boldsymbol{C}$ en cada punto del espacio que "actúa" sobre una masa que podemos colocar ahí, pero está allí, sea que realmente suministremos una masa para que "actúe" o no. $\boldsymbol{C}$ tiene tres componentes y cada una de esas componentes es una función de (x, y, z), una función de posición en el espacio. Tal cosa la llamamos *campo*, y decimos que dos objetos A y B *generan* un campo, esto es, "forman" el vector $\boldsymbol{C}$. Cuando se coloca un objeto en un campo, la fuerza sobre él es igual a su masa por el valor del vector campo en el punto en que se coloca el objeto.

También podemos hacer lo mismo con la energía potencial. Ya que la energía potencial, la integral de (−**fuerza**)·($d\boldsymbol{s}$), puede escribirse como m por la integral de (−**campo**)·($d\boldsymbol{s}$), un mero cambio de escala, vemos que la energía potencial $U(x, y, z)$ de un objeto situado en el punto (x, y, z) en el espacio, puede escribirse como m por otra función que podemos designar *potencial* Ψ. La integral $\int \boldsymbol{C} \cdot d\boldsymbol{s} = -\Psi$, como $\int \boldsymbol{F} \cdot d\boldsymbol{s} = -U$; hay sólo un factor de escala entre los dos:

$$U = -\int \boldsymbol{F} \cdot d\boldsymbol{s} = -m \int \boldsymbol{C} \cdot d\boldsymbol{s} = m\Psi. \tag{14.7}$$

Teniendo esta función $\Psi(x, y, z)$ en cada punto del espacio, inmediatamente podemos calcular la energía potencial de un objeto en cualquier punto del espacio, a saber, $U(x, y, z) = m\Psi(x, y, z)$ —parece un asunto fácil—. Pero en realidad no es trivial, porque a veces es mejor describir un campo dando el valor de Ψ en todas partes del espacio en vez de tener que dar $\boldsymbol{C}$. En vez de tener que escribir tres componentes complicadas de una función vectorial, podemos remplazarla por la función escalar Ψ. Además, es mucho más fácil calcular Ψ que cualquier componente dada de $\boldsymbol{C}$ cuando el campo es producido por cierto número de masas, porque dado que el potencial es escalar simplemente sumamos, sin preocuparnos de la dirección. Además, el campo $\boldsymbol{C}$ puede recuperarse fácilmente de Ψ como veremos luego. Supongan que tenemos masas puntuales m_1, $m_2 \ldots$ en los puntos $1, 2, \ldots$ y queremos saber el potencial Ψ en cierto punto arbitrario p. Esto es sencillamente la suma de los potenciales en p debido a las masas tomadas una por una:

$$\Psi(p) = \sum_i -\frac{Gm_i}{r_{ip}}, \qquad i = 1,\, 2,\, \ldots \tag{14.8}$$

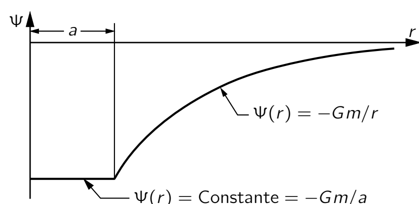


Figura 14-4. Potencial debido a una capa esférica de radio a.

En el capítulo anterior usamos esta fórmula, que el potencial es la suma de los potenciales de todos los diferentes objetos, para calcular el potencial debido a una capa esférica de materia sumamos las contribuciones al potencial en un punto de todas las partes de la capa. El resultado de este cálculo se muestra gráficamente en la figura 14-4. Es negativo, teniendo el valor cero cuando $r = \infty$ y variando con $1/r$ hasta el radio a y luego es constante en el interior de la capa. Fuera de la capa el potencial es $-Gm/r$, en que m es la masa de la capa, que es exactamente el mismo que habría sido si toda la masa estuviera ubicada en

el centro. Pero no es exactamente igual *en todas partes*, porque dentro de la capa el potencial resulta ser $-Gm/a$ y ¡es una constante! *Cuando el potencial es constante, no hay campo*, o cuando la energía potencial es constante no hay fuerza, porque si movemos un objeto de un lugar a otro cualquiera dentro de la esfera, el trabajo efectuado por la fuerza es exactamente cero. ¿Por qué? Porque el trabajo efectuado al mover el objeto de un lugar a otro es igual a menos el cambio de energía potencial (o bien la integral correspondiente del campo es la variación de potencial). Pero la energía potencial es la *misma* en cualquier par de puntos en el interior, de manera que la variación de energía potencial es cero y por consiguiente no se efectúa trabajo al moverse entre dos puntos dentro de la capa. La única manera de que el trabajo sea cero para todas las direcciones del desplazamiento, es que no haya ninguna fuerza.

Esto nos da una clave acerca de cómo podemos obtener la fuerza o el campo, dada la energía potencial. Supongamos que la energía potencial de un objeto se conoce en la posición (x, y, z) y deseamos saber cuál es la fuerza que actúa sobre el objeto. No basta conocer el potencial sólo en este punto único, como veremos; se requiere el conocimiento del potencial en los puntos vecinos también. ¿Por qué? ¿Cómo podemos calcular la componente x de la fuerza? (Por supuesto, si podemos hacer esto, también podremos encontrar las componentes y y z, así conoceremos la fuerza total.) Ahora bien, si moviéramos el objeto una distancia pequeña Δx, el trabajo efectuado por la fuerza que actúa sobre el objeto sería la componente x de la fuerza multiplicada por Δx, si Δx es lo suficientemente pequeño, y este producto será igual a la variación de energía potencial al ir de un punto al otro:

$$\Delta W = -\Delta U = F_x \, \Delta x. \tag{14.9}$$

Hemos usado simplemente la fórmula $\int \mathbf{F} \cdot ds = -\Delta U$, pero sólo para una trayectoria *muy corta*. Ahora dividimos entre Δx y encontramos que la fuerza es

$$F_x = -\Delta U/\Delta x. \tag{14.10}$$

Por cierto que esto no es exacto. Lo que en realidad deseamos es el límite de (14.10) cuando Δx se hace más y más pequeño porque es sólo *exactamente* correcta en el límite de Δx infinitesimal. Reconocemos esto como la derivada de U respecto a x, y nos inclinaríamos, por consiguiente, a escribir $-dU/dx$. Pero U depende de x, y y z, y los matemáticos han inventado un símbolo diferente para recordarnos que debemos tener mucho cuidado cuando derivamos una función como ésa, para recordar que estamos considerando que *sólo x varía* mientras que y y z no varían. En vez de una d sencillamente hacen "un 6 al revés" o ∂. (Un ∂ debiera haberse usado desde el comienzo del cálculo diferencial porque siempre queremos simplificar esa d, pero nunca queremos simplificar un ∂.) De manera que escriben a $\partial U/\partial x$, y además, en caso de apuro, si desean ser muy cuidadosos, colocan una línea al lado con una pequeña yz abajo ($\partial U/\partial x|_{yz}$), que significa "Tome la derivada de U respecto a x, manteniendo constantes y y z". Muy a menudo omitimos la observación sobre lo que se mantiene constante, porque generalmente es evidente del contexto, de manera que no usamos la línea con la y y la z. Sin embargo, *siempre* usamos un ∂ en vez de una d como advertencia de que es una derivada con algunas otras variables que se mantienen constantes. A esto se le llama *Derivada parcial*; es una derivada en que sólo variamos x.

Por consiguiente, encontramos que la fuerza que actúa en la dirección x es menos la derivada parcial de U respecto a x:

$$F_x = -\partial U/\partial x. \tag{14.11}$$

De manera similar, la fuerza en la dirección y puede encontrarse derivando U respecto a y, manteniendo x y z constantes, y la tercera componente, por supuesto, es la derivada respecto a z, manteniendo y y x constantes:

$$F_y = -\partial U/\partial y, \qquad F_z = -\partial U/\partial z. \tag{14.12}$$

Ésta es la manera de llegar de la energía potencial a la fuerza. Obtenemos el *campo* a partir del *potencial* de exactamente la misma manera:

$$C_x = -\partial\Psi/\partial x, \qquad C_y = -\partial\Psi/\partial y, \qquad C_z = -\partial\Psi/\partial z. \tag{14.13}$$

Entre paréntesis mencionaremos aquí otra notación, que no usaremos realmente por mucho tiempo: como $\mathbf{C}$ es un vector y tiene componentes x, y y z, los símbolos $\partial/\partial x$, $\partial/\partial y$ y $\partial/\partial x$ que producen las componentes x, y, z son algo así como vectores. Los matemáticos han inventado un magnífico nuevo símbolo, $\boldsymbol{\nabla}$, que se llama "grad" o "gradiente" que no es una cantidad sino un *operador*, que crea un vector a partir de un escalar. Tiene las siguientes "componentes": la componente x de este "grad" es $\partial/\partial x$, la componente y es $\partial/\partial y$ y la componente z es $\partial/\partial z$, y entonces tenemos el gusto de escribir nuestra fórmula de esta manera:

$$\boldsymbol{F} = -\boldsymbol{\nabla} U, \qquad \boldsymbol{C} = -\boldsymbol{\nabla}\Psi. \tag{14.14}$$

El usar $\boldsymbol{\nabla}$ nos da una manera rápida de probar si tenemos o no una verdadera ecuación vectorial, pero en realidad la ecuación (14.14) significa precisamente lo mismo que las ecuaciones (14.11), (14.12) y (14.13); es sólo otra manera de escribirlas, y como no queremos escribir tres ecuaciones cada vez, usamos $\boldsymbol{\nabla} U$ en su lugar.

Un ejemplo más de campos y potenciales tiene que ver con el caso eléctrico. En el caso de la electricidad, la fuerza que actúa sobre un objeto estático es la carga por el campo eléctrico: $\boldsymbol{F} = q\boldsymbol{E}$. (En general, por supuesto, la componente x de una fuerza en un problema de electricidad tiene también una parte que depende del campo magnético. Es fácil demostrar a partir de la ecuación (12.11) que la fuerza que actúa sobre una partícula debido a campos magnéticos es siempre perpendicular a su velocidad, y también perpendicular al campo. Como la fuerza debida al magnetismo sobre una carga móvil es perpendicular a la velocidad, el magnetismo *no efectúa trabajo* sobre la carga móvil, porque el movimiento es perpendicular a la fuerza. Por consiguiente, al calcular teoremas sobre energía cinética en campos eléctricos y magnéticos, podemos omitir la contribución del campo magnético, ya que no hace variar la energía cinética.) Suponemos que existe sólo un campo eléctrico. Entonces podemos calcular la energía o el trabajo efectuado de la misma manera que para la gravedad y calcular una cantidad ϕ que es menos la integral de $\boldsymbol{E}\cdot d\boldsymbol{s}$, desde el punto fijo arbitrario al punto donde hacemos el cálculo, y entonces la energía potencial en un campo eléctrico es justamente la carga por esta cantidad ϕ.

$$\phi(\boldsymbol{r}) = -\int \boldsymbol{E}\cdot d\boldsymbol{s},$$

$$U = q\phi.$$

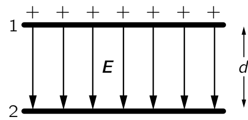


Figura 14-5. Campo entre dos placas paralelas.

Tomemos, como ejemplo, el caso de dos placas metálicas paralelas, cada una con carga superficial $\pm\sigma$ por unidad de área. Éste se llama condensador de placas paralelas. Encontramos anteriormente que la fuerza es cero fuera de las placas y que hay un campo eléctrico constante entre ellas, dirigido de $+$ a $-$ y de módulo σ/ϵ_0 (figura 14.5). Nos gustaría saber cuánto trabajo se efectuaría al llevar una carga de una placa a la otra. El trabajo sería la integral de (**fuerza**) $\cdot$ ($d\boldsymbol{s}$), que puede escribirse como carga por valor del potencial en la placa 1 menos el de la placa 2:

$$W = \int_1^2 \boldsymbol{F}\cdot d\boldsymbol{s} = q(\phi_1 - \phi_2).$$

En realidad podemos calcular la integral, porque la fuerza es constante y si designamos por d la separación de las placas, la integral es fácil:

$$\int_1^2 \boldsymbol{F}\cdot d\boldsymbol{s} = \frac{q\sigma}{\epsilon_0}\int_1^2 dx = \frac{q\sigma d}{\epsilon_0}.$$

La diferencia de potencial, $\Delta\phi = \sigma d/\epsilon_0$, se llama *diferencia de voltaje*, y ϕ se mide en volts. Cuando decimos que un par de placas están cargadas a cierto voltaje, lo que queremos decir es que la diferencia de potencial eléctrico de las dos placas es de tantos voltios. Para un capacitor hecho de dos placas paralelas que llevan una carga superficial $\pm\sigma$, el voltaje o diferencia de potencial del par de placas es $\sigma d/\epsilon_0$.

15

Teoría especial de la relatividad

15-1 El principio de relatividad

Por más de 200 años se pensó que las ecuaciones del movimiento formuladas por Newton describían correctamente la naturaleza y cuando por primera vez se descubrió un error en estas leyes, también se descubrió la manera de corregirlo. Tanto el error como su corrección fueron descubiertos por Einstein en 1905.

La segunda ley de Newton que hemos expresado por medio de la ecuación

$$F = d(mv)/dt,$$

fue establecida con la hipótesis tácita que m es una constante, pero sabemos ahora que esto no es cierto y que la masa de un cuerpo aumenta con su velocidad. En la fórmula corregida de Einstein, m tiene el valor

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \tag{15.1}$$

donde la "masa en reposo" m_0 representa la masa de un cuerpo que no se mueve y c es la velocidad de la luz, que es alrededor de 3×10^5 km $\cdot$ s $^{-1}$, o sea, alrededor de $300\,000\,\text{km} \cdot \text{s}^{-1}$.

Para aquellos que quieren aprender sólo lo suficiente en esta materia para poder resolver problemas, esto es todo lo que hay que saber sobre la teoría de la relatividad —se cambian las leyes de Newton introduciendo un factor de corrección para la masa—. De la misma fórmula se puede ver fácilmente que este aumento de masa es muy pequeño en circunstancias normales. Incluso para velocidades tan grandes como las de un satélite que se mueve alrededor de la Tierra con 8 km/s, se tiene $v/c = 8/300\,000$ es igual a introducir este valor en la fórmula y se ve que la corrección a la masa es solamente una parte en 2 000 a 3 000 millones que es casi imposible de observar. En realidad, la exactitud de la fórmula ha sido confirmada ampliamente observando muchos tipos de partículas que se mueven con velocidades hasta prácticamente la Velocidad des luz. Sin embargo, debido a que el efecto es normalmente tan pequeño, resulta notable que haya sido descubierto teóricamente antes que experimentalmente. De forma empírica, el efecto es muy grande a velocidades suficientemente elevadas, pero no fue descubierto de esta manera. Por esta razón es interesante ver cómo una ley que implicaba una modificación tan delicada (cuando fue descubierta por primera vez) se encontró por medio de una combinación de experimentos y razonamientos físicos. Contribuciones al descubrimiento fueron hechas por muchas personas cuyo resultado final fue el descubrimiento de Einstein.

En realidad existen dos teorías de la relatividad de Einstein. Este capítulo solamente está dedicado a la teoría especial de relatividad que fue formulada en 1905. En 1915 Einstein publicó una teoría adicional llamada teoría general de la relatividad. Esta última teoría da una extensión de la teoría especial al caso de la ley de gravitación, no discutiremos aquí la teoría general.

El principio de relatividad fue formulado por primera vez por Newton en sus corolarios de las leyes de movimiento: "Los movimientos de cuerpos en un mismo espacio dado son iguales entre sí, si este espacio está en reposo o si se mueve uniformemente sobre una línea recta". Esto significa, por ejemplo, que si una nave espacial se desplaza con una velocidad uniforme, todos los experimentos y los fenómenos realizados en ella aparecerán iguales a los observados si la nave no está en movimiento, suponiendo, naturalmente, que uno no mira hacia fuera.

Éste es el significado del principio de relatividad. La idea es bastante sencilla, y la única pregunta es si es *verdad* que en todos los experimentos realizados en el interior de un sistema en movimiento las leyes de la física aparecen iguales a las observadas si el sistema está en reposo. Investiguemos primero si las leyes de Newton son iguales en el sistema móvil. Supongan que Pedro se está moviendo en la dirección x con velocidad uniforme u y que mide la posición de un cierto punto, indicado en la figura 15-1. Él designa la "distancia x" del punto en su sistema de coordenadas con x', Juan está en reposo, y mide la distancia del mismo punto, designando la coordenada x en su sistema con x. La relación entre las coordenadas en los dos sistemas se ve claramente en el diagrama. Después de un tiempo t el origen de Pedro se ha movido una distancia ut, y si originalmente los dos sistemas coincidieron,

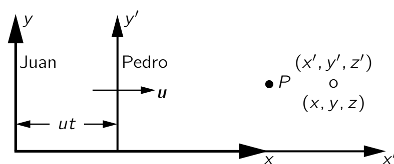


Figura 15-1. Dos sistemas de coordenadas en movimiento uniforme relativo según sus ejes x.

$$
\begin{aligned}
x' &= x - ut,\\
y' &= y,\\
z' &= z,\\
t' &= t.
\end{aligned}
\tag{15.2}
$$

Si sustituimos esta transformación de coordenadas en las leyes de Newton encontramos que estas leyes se transforman en las mismas leyes en el sistema primado, es decir, las leyes de Newton tienen las mismas formas en un sistema en movimiento que en un sistema estático y por esta razón es imposible poder decir por medio de experimentos mecánicos si el sistema se está moviendo o no.

El principio de relatividad se ha usado en mecánica durante largo tiempo. Fue usado por diferentes personas, en particular por Christiaan Huygens, para obtener las reglas de la colisión de bolas de billar prácticamente en la misma forma usada en el capítulo 10 para discutir la conservación del momentum. En el siglo XIX aumentó el interés en aquel principio como resultado de las investigaciones de los fenómenos de electricidad, magnetismo y luz. Una larga serie de experimentos cuidadosos acerca de estos fenómenos realizados por muchas personas, culminó en las ecuaciones de Maxwell para el campo electromagnético, las cuales describen la electricidad, el magnetismo y la luz en un único sistema uniforme. Sin embargo, las ecuaciones de Maxwell *no* parecían obedecer el principio de relatividad. Es decir, si transformamos las ecuaciones de Maxwell por medio de la sustitución de las ecuaciones (15.2), *su forma no queda igual*, por esta razón, los fenómenos eléctricos y ópticos en una nave espacial en movimiento deberían ser diferentes a los mismos en una nave en reposo. Se podrían usar entonces estos fenómenos ópticos para determinar la velocidad de la nave; en particular, se podría determinar la velocidad absoluta de la nave, efectuando mediciones ópticas o eléctricas adecuadas. Una de las consecuencias de las ecuaciones de Maxwell es que, si existe una perturbación en el campo tal que se genera luz, estas ondas electromagnéticas se alejan en todas las direcciones en la misma forma y con la misma velocidad c, o sea 300 000 km/s. Otra consecuencia de las ecuaciones es que si la fuente de la perturbación se mueve, la luz emitida atraviesa el espacio con la misma velocidad c. Esto es análogo al caso del sonido, donde la velocidad de las ondas sonoras es también independiente de la velocidad de la fuente.

Esta independencia del movimiento de la fuente, en el caso de la luz, nos plantea un problema interesante:

Supongan que nos encontramos en un auto que se mueve con la velocidad u, y que la luz proveniente de su parte trasera y pasa por el automóvil con velocidad c. Derivando la primera ecuación en (15.2) se obtiene:

$$dx'/dt = dx/dt - u,$$

lo que significa que, según la transformación de Galileo, la velocidad aparente de la luz que pasa, medida en el automóvil, no debería ser c sino $c - u$. Por ejemplo, si el automóvil se mueve a 200 000 km/s, y si la luz se mueve a 300 000 km/s, entonces aparentemente la luz que pasa por el automóvil debería desplazarse

a 100 000 km/s. Por lo tanto, midiendo la velocidad de la luz que pasa por el automóvil (si la transformación de Galileo es correcta para la luz), se puede determinar la velocidad del automóvil. Se realizaron varios experimentos basados en esta idea general, para determinar la velocidad de la Tierra, pero todos fallaron —*no dieron ninguna velocidad*—. Discutiremos uno de estos experimentos en detalle, para mostrar exactamente qué se hizo y qué *pasó*, naturalmente algo pasó, algo falló con las ecuaciones de la física. ¿Qué podría ser?

15-2 La transformación de Lorentz

Cuando se descubrió el fracaso de las ecuaciones de la física en el caso citado arriba, lo primero que se pensó era que la dificultad debía estar en las nuevas ecuaciones de Maxwell de la electrodinámica, que en esos tiempos sólo tenían 20 años. Parecía evidente que estas ecuaciones debían estar equivocadas, por lo cual había que cambiarlas de tal manera que, en una transformación de Galileo, se satisficiera el principio de relatividad. Cuando se intentó esto, los nuevos términos que había que colocar en las ecuaciones conducían a predicciones de nuevos fenómenos eléctricos que no existían al ser buscados experimentalmente, por lo cual este intento fue abandonado. Gradualmente se reconoció entonces que las leyes de Maxwell de la electrodinámica eran correctas y que los problemas debían ser buscados en otra parte.

Mientras tanto, Hendrik Antoon Lorentz observó algo notable y curioso al efectuar las siguientes sustituciones en las ecuaciones de Maxwell:

$$\begin{aligned} x' &= \frac{x - ut}{\sqrt{1 - u^2/c^2}}, \\ y' &= y, \\ z' &= z, \\ t' &= \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}}, \end{aligned} \tag{15.3}$$

es decir, que las ecuaciones de Maxwell mantienen la misma forma cuando se les aplica esta transformación. Las ecuaciones (15.3) son conocidas bajo el nombre de *transformación de Lorentz*. Siguiendo una sugerencia hecha originalmente por Henri Poincaré, Einstein propuso que *todas las leyes físicas* debían ser tales de *permanecer las mismas bajo una transformación de Lorentz*. En otras palabras, deberíamos cambiar las leyes de la mecánica y no las leyes de la electrodinámica. ¿Cómo debemos cambiar las leyes de Newton para que *queden* inalteradas por la transformación de Lorentz? Si nos fijamos esta meta, tendremos que reescribir las ecuaciones de Newton de tal manera que las condiciones impuestas sean satisfechas. Se obtiene así que el único requisito es que la masa m en las ecuaciones de Newton debe ser remplazada por la expresión indicada en la ecuación (15.1). Si se efectúa este cambio, armonizan las leyes de Newton y las leyes de la electrodinámica. Si usamos entonces la transformación de Lorentz para comparar las mediciones de Pedro con las de Juan, no podremos detectar nunca quién se está moviendo, porque la forma de todas las ecuaciones será la misma en ambos sistemas de coordenadas.

Es interesante discutir lo que significa el remplazo de la transformación antigua para las coordenadas y el tiempo por una nueva, dado que la antigua (Galileo) parece evidente, y la nueva (Lorentz) tiene un aspecto tan peculiar. Deseamos saber, si es lógico y experimentalmente posible, que la nueva transformación, y no la antigua, sea la correcta. Para saber esto, no basta estudiar las leyes de la mecánica, sino que, como lo hizo Einstein, debemos analizar también nuestros conceptos sobre *espacio* y *tiempo* para poder comprender esta transformación. Tendremos que discutir con algún detalle estos conceptos y sus consecuencias para la mecánica, digamos mientras tanto que el esfuerzo será justificado, dado que los resultados están de acuerdo con la experiencia.

15-3 El experimento de Michelson-Morley

Como se mencionó antes, se hicieron intentos para determinar la velocidad absoluta de la Tierra a través del hipotético "éter", el cual supuestamente se extiende por todo el espacio. El más famoso de estos experimentos de Stern-Gerlachs es uno realizado por Michelson y Morley en 1887. Sólo 18 años más tarde Einstein finalmente explicó los resultados negativos del experimento.

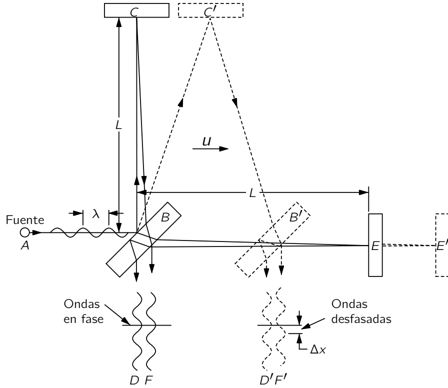


Figura 15-2. Diagrama esquemático del experimento de Michelson-Morley.

El experimento de Michelson-Morley fue realizado con un aparato como el que se muestra esquemáticamente en la figura 15-2. Este aparato consta esencialmente de una fuente luminosa A, una placa de vidrio parcialmente plateada B y dos espejos C y E, todo montado sobre una base rígida. Los espejos están puestos a iguales distancias L de B. La placa B divide el haz de la luz incidente y los dos haces resultantes siguen en direcciones perpendiculares entre sí hacia los espejos, donde se reflejan de vuelta hacia B. Al volver a B, los dos haces se recombinan, formando los haces superpuestos D y F. Si el tiempo que emplea la luz para avanzar de B a E y viceversa es el mismo que el tiempo B a C y viceversa, entonces los haces D y F que emergen estarán en fase y se reforzarán mutuamente, pero si los dos tiempos difieren ligeramente, los haces estarán ligeramente fuera de fase y se obtendrá interferencia. Si el aparato está "en reposo" en el éter, los tiempos deberán ser exactamente iguales, pero si se mueve hacia la derecha con velocidad u debe haber una diferencia en los tiempos. Veamos por qué.

Calculemos primero el tiempo que necesita la luz para ir de B a E y viceversa. Llamemos t_1 el tiempo que requiere la luz para ir de la placa B al espejo E y t_2 el tiempo de regreso. Ahora bien, mientras la luz se mueve de B hacia el espejo, el aparato se mueve a una distancia ut_1, tal que la luz debe atravesar una distancia $L + ut_1$, con la velocidad c. Podemos también expresar esta distancia como ct_1, así tenemos:

$$ct_1 = L + ut_1, \qquad \text{o} \qquad t_1 = L/(c - u).$$

(Este resultado es también evidente bajo el punto de vista de que la velocidad de la luz respecto al aparato es $c - u$, de modo que el tiempo es la longitud L dividido entre $c - u$.) En una forma semejante se puede calcular el tiempo t_2. Durante este tiempo la placa B avanza una distancia ut_2; por lo tanto, la distancia de regreso de la luz es $L - ut_2$. Tenemos, en consecuencia,

$$ct_2 = L - ut_2, \qquad \text{o} \qquad t_2 = L/(c + u).$$

Entonces el tiempo total es

$$t_1 + t_2 = 2Lc/(c^2 - u^2).$$

Por conveniencia, en comparaciones posteriores de tiempo, escribimos esto en la forma

$$t_1 + t_2 = \frac{2L/c}{1 - u^2/c^2}. \tag{15.4}$$

Nuestro segundo cálculo será el del tiempo t_3 que emplea la luz para ir de B al espejo C. Como en el caso anterior, durante el tiempo t_3 el espejo se mueve hacia la derecha una distancia ut_3; hasta la posición C', en el mismo tiempo la luz viaja una distancia ct_3 a lo largo de la hipotenusa de un triángulo que es BC'. Para este triángulo rectángulo tenemos:

$$(ct_3)^2 = L^2 + (ut_3)^2$$

o bien

$$L^2 = c^2t_3^2 - u^2t_3^2 = (c^2 - u^2)t_3^2,$$

de lo cual obtenemos

$$t_3 = L/\sqrt{c^2 - u^2}.$$

Para el viaje de regreso desde C' la distancia es la misma, como puede deducirse de la simetría de la figura; por esta razón, también el tiempo de regreso

es el mismo, y el tiempo total es $2t_3$. Con un pequeño reordenamiento de la forma, podemos escribir:

$$2t_3 = \frac{2L}{\sqrt{c^2 - u^2}} = \frac{2L/c}{\sqrt{1 - u^2/c^2}}. \tag{15.5}$$

Estamos ahora en condiciones de comparar los tiempos empleados por los dos haces de luz. En las expresiones (15.4) y (15.5) los numeradores son iguales y representan el tiempo que se emplearía si el aparato estuviera en reposo. En los denominadores el término u_2/c_2 será pequeño, salvo que u sea de valor comparable a c. Los denominadores representan las modificaciones en los tiempos causados por el movimiento del aparato. Y vean, estas modificaciones *no son iguales:* el tiempo para ir a C y viceversa es un poco más pequeño que el tiempo a E y viceversa, a pesar de que los espejos están equidistantes de B, y todo lo que tenemos que hacer es medir esta diferencia con precisión.

Aquí surge un aspecto menor de tipo técnico —supóngase que las dos longitudes L no son exactamente iguales—. De hecho es seguro que no las podemos hacer exactamente iguales. En este caso simplemente giramos el aparato en 90 grados, tal que BC esté en la línea de movimiento y BE perpendicular al movimiento. Cualquier pequeña diferencia de la longitud no tiene entonces importancia, y lo que observamos es un *corrimiento* de las franjas de interferencia cuando se gira el aparato.

Al efectuar el experimento, Michelson y Morley orientaron el aparato de tal manera que la línea BE quedó casi paralela al movimiento de la Tierra en su órbita (durante ciertos momentos del día o de la noche). Esta velocidad orbital es aproximadamente 30 kilómetros por segundo y cualquier "desplazamiento del éter" debería ser por lo menos éste en algún momento del día o de la noche y en alguna época del año. El aparato fue lo suficientemente sensible para observar tal efecto, pero no se encontró ninguna diferencia de tiempo —la velocidad de la Tierra a través del éter no pudo ser detectada—. El resultado del experimento fue cero.

El resultado del experimento de Michelson-Morley fue muy enigmático y desconcertante. La primera idea fructífera para encontrar una salida a este *impasse* vino de Lorentz. Sugirió que cuerpos materiales se contraen cuando se mueven, y que este acortamiento tiene lugar solamente en la dirección del movimiento, y además, que si el largo es L_0 cuando el cuerpo está en reposo, entonces si se mueve con velocidad u paralela a su largo, este nuevo largo que llamaremos $L_\parallel$ (L paralelo) estará dado por:

$$L_\parallel = L_0\sqrt{1 - u^2/c^2}. \tag{15.6}$$

Si se aplica esta modificación al interferómetro de Michelson-Morley no cambia la distancia de B a C, pero la distancia de B a E se acorta a $L/\sqrt{1 - u^2/c^2}$. Por esta razón la ecuación (15.5) no cambia, pero la L de la ecuación (15.4) se debe cambiar de acuerdo con la ecuación (15.6). Al hacer esto obtenemos:

$$t_1 + t_2 = \frac{(2L/c)\sqrt{1 - u^2/c^2}}{1 - u^2/c^2} = \frac{2L/c}{\sqrt{1 - u^2/c^2}}. \tag{15.7}$$

Comparando este resultado con la ecuación (15.5) vemos que $t_1 + t_2 = 2t_3$. Así que si el aparato se contrae en la forma recién descrita tenemos una manera de comprender por qué el experimento de Michelson-Morley no da efecto alguno. A pesar de que la hipótesis de contracción explica el resultado negativo del experimento, queda la objeción que fue inventada expresamente para explicar esta dificultad y que es demasiado artificiosa. Sin embargo, en muchos otros experimentos encaminados a descubrir un viento de éter, aparecieron dificultades similares, hasta que fue aparente que la naturaleza estaba en una "conspiración" para frustrar al hombre, introduciendo algunos fenómenos nuevos para anular todo fenómeno que él creía le iba a permitir una medición de u.

Finalmente se reconoció, como lo hizo ver Poincaré, que una *¡conspiración total es de por sí una ley de la naturaleza!* Entonces, Poincaré propuso que *existe*

tal ley de la naturaleza y que es imposible detectar un viento de éter con *ningún* experimento, es decir, no existe manera para determinar una velocidad absoluta.

15-4 Transformación del tiempo

Al tratar de probar si la idea de la contracción estaba en armonía con los hechos de otros experimentos, resultó que todo era correcto si también se modificaban los *tiempos* en la forma expresada por la cuarta ecuación del grupo (15.3). Esto se debe a que el tiempo t_3 calculado para el viaje de B a C y viceversa no es el mismo si lo calcula un hombre que realiza el experimento en una nave espacial en movimiento, a si lo calcula un observador estático, quien observa a la nave espacial. Para el hombre en la nave, el tiempo es simplemente $2L/c$, sin embargo para el otro observador es $(2L/c)/\sqrt{1-u^2/c^2}$ (ecuación (15.5)). En otras palabras, si el observador que está fuera de la nave ve que el hombre en la nave espacial enciende un cigarro, todas las acciones aparecen más lentas de lo normal, mientras que para el hombre en el interior todo se mueve a una velocidad normal. Por lo tanto, no solamente las longitudes deben ser reducidas, sino que también los instrumentos que miden tiempos (relojes) aparentemente deben ir más lentos. Es decir, si el reloj en la nave espacial registra que para el hombre en la nave ha transcurrido un segundo, indicará $1/\sqrt{1-u^2/c^2}$ segundos para el hombre afuera.

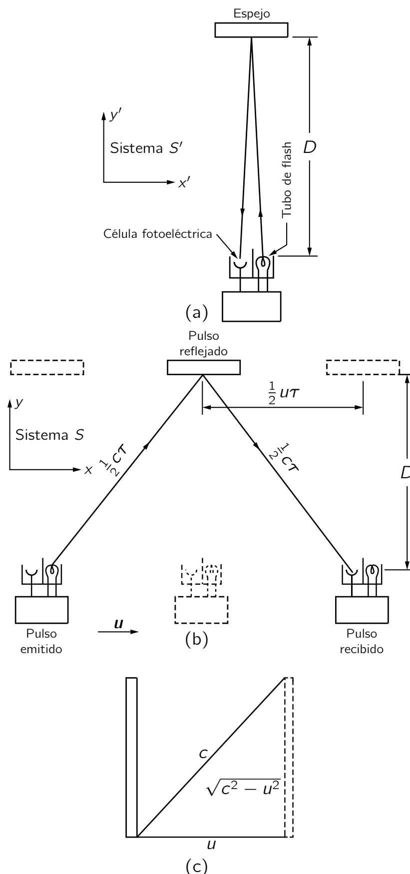


Figura 15-3. (a) Un "reloj de luz" en reposo en el sistema S'. (b) El mismo reloj en movimiento a través del sistema S. (c) Ilustración del camino diagonal recorrido por el haz de luz en un "reloj de luz" en movimiento.

Este atraso de los relojes en un sistema en movimiento es un fenómeno muy singular y es digno de una explicación. Para poder entenderla, debemos observar el mecanismo del reloj y ver qué pasa cuando se mueve. Dado que esto es algo difícil, vamos a considerar un reloj muy sencillo. El que vamos a elegir es ciertamente un tipo de reloj muy tonto, pero trabajará en principio: consiste de una varilla (metro) con un espejo en cada extremo, y si emitimos una señal luminosa entre los espejos, la luz sigue moviéndose hacia arriba y hacia abajo, produciendo un clic cada vez que llegue abajo, como un reloj normal. Construimos dos relojes de esta forma, con exactamente la misma longitud, y los sincronizamos haciéndolos partir al mismo tiempo, entonces concordarán siempre, dado que son iguales en longitud, y la luz viaja siempre con velocidad c. Entregamos uno de estos relojes al hombre que lo llevará en su nave espacial instalando la varilla perpendicular a la dirección del movimiento de la nave, entonces no cambiará el largo de la varilla. ¿Cómo sabemos que los largos perpendiculares no cambian? Los dos hombres pueden acordar hacer marcas en cada uno de sus metros en la dirección y cuando pasan uno frente al otro. Por simetría, las dos marcas deben aparecer en las mismas coordenadas y y y', porque de otra manera cuando se juntan para comparar resultados, una marca estará por encima o por debajo de la otra, y podríamos decir así quién realmente se estaba moviendo.

Veamos ahora qué le pasa al reloj en movimiento. Antes de llevarlo a bordo el hombre estaba de acuerdo en que era un hermoso reloj patrón y cuando se aleja en la nave espacial, no notará nada peculiar. Si lo hiciera, podría saber que estaba en movimiento, si cualquier cosa cambiara debido al movimiento, él podría decir que estaba en movimiento. Pero el principio de relatividad afirma que esto es imposible en un sistema que se mueve uniformemente; por lo tanto, nada ha cambiado. Por otra parte, si el observador externo mira el reloj que pasa frente a él, ve que la luz, al ir de espejo a espejo, recorre "en realidad" un camino en zigzag, dado que la varilla se mueve lateralmente todo el tiempo. Hemos analizado ya tal movimiento en zigzag en conexión con el experimento de Michelson-Morley. Si en un tiempo dado la varilla se mueve una distancia proporcional a u (figura 15-3), la distancia que la luz recorre en el mismo tiempo es proporcional a c, y la distancia vertical es por lo tanto proporcional a $\sqrt{c^2-u^2}$.

O sea, la luz requiere un *tiempo mayor* para ir de extremo a extremo en el reloj en movimiento que en el reloj en reposo. Por esta razón, el tiempo aparente entre los clic es mayor para el reloj en movimiento, en la misma proporción que la indicada por la hipotenusa del triángulo (éste es también el origen de las raíces cuadradas en nuestras ecuaciones). En la figura también es evidente que mientras más grande es u, más lentamente parece marchar el reloj en movimiento. No

sólo se mueve mas lentamente este reloj particular, sino que, si la teoría de la relatividad es correcta, cualquier otro reloj que funcione según cualquier principio también parecería marchar lentamente, y en la misma proporción podemos decir esto sin más análisis. ¿Por qué es así?

Para contestar la pregunta hecha arriba, supongan que tuviéramos otros dos relojes, hechos exactamente iguales, con ruedas y engranajes, o quizá basados en la desintegración radiactiva o cualquier otra cosa. Entonces ajustamos estos relojes para que ambos marchen en perfecto sincronismo con nuestros relojes originales. Si la luz camina hacia arriba y hacia abajo en los primeros relojes y anuncia su llegada con un clic, los modelos nuevos también completan alguna especie de ciclo, lo que ambos anuncian simultáneamente por algún destello doblemente coincidente, o un bum, o cualquier otra señal. Uno de estos relojes se lleva a la nave espacial junto con uno del primer tipo. A lo mejor *este* reloj no marcha más lentamente, sino que sigue manteniendo el mismo tiempo que su contraparte en reposo y así discrepa con el otro reloj en movimiento. Ah, no, si esto pasara, el hombre en la nave podría usar esta diferencia entre los dos relojes para determinar la velocidad de su nave; lo que hemos supuesto es imposible. *No necesitamos saber nada sobre el mecanismo* del reloj nuevo que pueda producir el efecto; sabemos simplemente que, cualquiera que sea la razón, aparentemente marchará más lento, igual que el primero.

Ahora bien, si *todos* los relojes en movimiento marchan más lentamente, si ningún método de medida del tiempo da otra cosa que una velocidad más baja, no nos queda otra cosa que decir, en cierto sentido, que el *tiempo mismo* parece más lento en una nave espacial. Todos los fenómenos que ocurren allí —el ritmo del pulso del hombre, sus procesos mentales, el tiempo que emplea para encender un cigarro, cuánto necesita para crecer y envejecer—, todas estas cosas deben ser más lentas en la misma proporción, porque él no puede decir que está en movimiento. Los biólogos y médicos dicen a veces que no es muy cierto que el tiempo necesario para el desarrollo de un cáncer es mayor en una nave espacial, pero desde el punto de vista de un físico moderno esto es casi cierto, ¡de otra manera se podría usar la velocidad de desarrollo del cáncer para determinar la velocidad de la nave!

Un ejemplo muy interesante de la dilatación del tiempo con el movimiento es suministrado por los muones, que son partículas que se desintegran espontáneamente después de un tiempo de vida medio de 2.2×10^{-6} s. Llegan a la Tierra en los rayos cósmicos, pero también pueden ser producidos artificialmente en el laboratorio. Algunos se desintegran en medio del aire, pero el resto se desintegra solamente después de haber encontrado un pedazo de material y haberse detenido. Está claro que en su vida tan corta el muon no puede viajar mucho más de 600 metros, incluso a la velocidad de la luz. Sin embargo, a pesar de que los muones se forman en la parte superior de la atmósfera, a unos 10 kilómetros de altura se les encuentra en el laboratorio aquí abajo en los rayos cósmicos. ¿Cómo puede ser esto? La contestación es que los diferentes muones se mueven con varias velocidades, algunas de las cuales son muy cercanas a la velocidad de la luz. Mientras que desde su propio punto de vista viven solamente 2μs, desde nuestro punto de vista viven considerablemente más —lo suficiente para que puedan llegar a la Tierra—. El factor por el cual se aumenta el tiempo ya ha sido dado como $1/\sqrt{1-u^2/c^2}$. La vida media ha sido medida con bastante exactitud para muones a diferentes velocidades, y los valores concuerdan bastante bien con la fórmula.

No sabemos por qué el muon se desintegra y cuál es su mecanismo, pero sabemos que su comportamiento satisface al principio de relatividad. Esta es la utilidad del principio de relatividad, nos permite hacer predicciones, incluso sobre cosas de las cuales no sabemos mucho en otro aspecto. Por ejemplo, antes de tener alguna idea sobre lo que hace desintegrar al muon, podemos predecir que si se mueve a nueve décimas de la velocidad de la luz, la duración aparente de su vida es $(2.2 \times 10^{-6})/\sqrt{1-9^2/10^2}$ s, y nuestra predicción funciona —esto es lo bueno—.

15-5 La contracción de Lorentz

Volvamos ahora a la transformación de Lorentz (15.3) y tratemos de entender mejor la relación entre los sistemas de coordenadas (x, y, z, t) y (x', y', z', t') a los que llamamos sistemas S y S', o sistemas de Juan y Pedro, respectivamente. Hemos observado ya que la primera de las ecuaciones está basada en la sugerencia de Lorentz de una contracción a lo largo de la dirección x, ¿cómo podemos demostrar que se produce una contracción? Ahora podemos darnos cuenta de que debido al principio de relatividad, el brazo *transversal* BC no puede cambiar su longitud en el experimento de Michelson-Morley; sin embargo, el resultado nulo del experimento exige que los *tiempos* sean iguales. Entonces, para que el experimento dé un resultado nulo, el brazo longitudinal BE debe aparecer $\sqrt{1-u^2/c^2}$ veces más corto. ¿Qué significa esta contracción en relación con las mediciones efectuadas por Juan y Pedro? Supongan que Pedro, que se mueve con el sistema S' en la dirección x, mide la coordenada x' de algún punto con un metro. Él coloca la varilla x' veces, pensando así que la distancia es x' metros. Desde el punto de vista de Juan en el sistema S, sin embargo, Pedro está usando una regla acortada tal que la distancia "real" medida es $x'\sqrt{1-u^2/c^2}$ metros. Entonces, si el sistema S' se ha alejado una distancia ut del sistema S, el observador S diría que el mismo punto, medido en sus coordenadas, está a una distancia $x = x'\sqrt{1-u^2/c^2} + ut$, o bien

$$x' = \frac{x - ut}{\sqrt{1-u^2/c^2}},$$

que es la primera ecuación de la transformación de Lorentz.

15-6 Simultaneidad

De manera análoga, debido a la diferencia en las escalas de tiempo, la expresión del denominador se ha introducido en la cuarta ecuación de la transformación de Lorentz. El término más interesante en esta ecuación es ux/c^2 en el numerador, porque es nuevo e inesperado. Ahora bien, ¿qué significa esto? Si observamos la situación cuidadosamente vemos que sucesos que ocurren al mismo tiempo en dos lugares separados vistos por Pedro en S', no ocurren al mismo tiempo al ser observados por Juan en S. Si un suceso ocurre en el punto x_1, en el tiempo t_0 y el otro suceso en x_2 y t_0 (al mismo tiempo), encontramos que los tiempos correspondientes t_1' y t_2' difieren en la cantidad

$$t_2' - t_1' = \frac{u(x_1 - x_2)/c^2}{\sqrt{1-u^2/c^2}}.$$

A este hecho se le llama "falta de simultaneidad a distancia", y para aclarar un poco más la idea consideremos el experimento siguiente.

Supongan que un hombre que se mueve en una nave espacial (sistema S') ha puesto un reloj en cada extremo de la nave y está interesado en asegurarse de que los dos relojes estén sincronizados. ¿Cómo se puede sincronizar los relojes? Existen muchos caminos. Un camino que implica muy poco cálculo sería localizar primeramente el punto medio exacto entre los relojes. Luego emitimos desde allí una señal luminosa que irá en ambas direcciones con la misma velocidad y evidentemente llegará a ambos relojes al mismo tiempo. Esta llegada simultánea de las señales se puede usar para sincronizar los relojes. Supongamos entonces que el hombre en S' sincroniza sus relojes por medio de este método particular. Veamos si un observador en el sistema S estaría de acuerdo que los dos relojes están sincronizados. El hombre en S' tiene derecho a pensar que lo están, porque no sabe que se está moviendo. Pero el hombre en S razona que, dado que la nave se está moviendo hacia adelante, el reloj en la parte delantera se aleja de la señal luminosa; por lo tanto, la luz tiene que viajar más que el medio camino para alcanzarlo; en cambio, el reloj trasero avanza para encontrar a la señal luminosa, por lo cual la distancia será más corta. La señal llega entonces primero al reloj trasero, a pesar de que el hombre en S' pensaba que ambas señales habían

llegado simultáneamente. Vemos entonces que si un hombre en una nave espacial cree que los tiempos en dos posiciones son simultáneos, ¡valores iguales de t' en su sistema de coordenadas deben corresponder a valores *diferentes* de t en otro sistema de coordenadas!

15-7 Vectores de cuatro dimensiones

Veamos qué otra cosa podemos descubrir en la transformación de Lorentz. Es interesante notar que la transformación entre los x y los t es de forma análoga a la transformación de los x y los y que estudiamos en el capítulo 11 para una rotación de coordenadas. Teníamos entonces

$$\begin{aligned} x' &= x\cos\theta + y\operatorname{sen}\theta, \\ y' &= y\cos\theta - x\operatorname{sen}\theta, \end{aligned} \tag{15.8}$$

donde los nuevos x' mezclan los antiguos x y y, y los nuevos y' también mezclan los antiguos x y y; en forma similar encontramos en la transformación de Lorentz un nuevo x' que es una mezcla de x y t y un nuevo t' que es una mezcla de t y x. Así, pues, la transformación de Lorentz es análoga a una rotación, sólo que es una "rotación" en el *espacio* y el *tiempo*, lo que parece ser un concepto extraño. Una prueba de la analogía con una rotación se puede hacer calculando la cantidad:

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = x^2 + y^2 + z^2 - c^2t^2. \tag{15.9}$$

En esta ecuación, los tres primeros términos de cada lado representan, en la geometría tridimensional, el cuadrado de la distancia entre un punto y el origen (superficie de una esfera) que no cambia (es invariante) en la rotación de los ejes coordenados. En forma similar, la ecuación (15.9) demuestra que existe una cierta combinación que incluye el tiempo, que es invariante en una transformación de Lorentz. Por lo tanto, la analogía con una rotación es completa, y es tal que vectores, es decir, cantidades con "componentes" que se transforman en la misma forma que las coordenadas y el tiempo, también son útiles en relatividad.

Consideremos entonces una extensión del concepto de vector, que hasta ahora sólo hemos considerado con componentes espaciales, para incluir una componente temporal. Es decir, esperamos que existan vectores con cuatro componentes, tres de las cuales son como las componentes de un vector ordinario, y a éstas se agregará una cuarta componente, que es análoga de la parte temporal.

Este concepto se analizará más en los capítulos siguientes, donde vamos a ver que si las ideas del párrafo anterior son aplicadas al momentum, la transformación da lugar a tres partes espaciales que son semejantes a las componentes del momentum ordinario, y una cuarta componente, la parte del tiempo, que es la *energía*.

15-8 Dinámica relativista

Estamos ahora en condiciones de investigar de manera más general qué forma toman las leyes de la mecánica frente a una transformación de Lorentz. [Hasta ahora hemos explicado cómo cambian la longitud y el tiempo, pero no cómo obtenemos la fórmula modificada para m, ecuación (15.1). Haremos esto en el próximo capítulo.] Para ver las consecuencias de la modificación de Einstein para la m de la mecánica newtoniana, empezaremos con la ley de Newton que establece que la fuerza es igual a la variación del momentun, es decir,

$$\boldsymbol{F} = d(m\boldsymbol{v})/dt.$$

El momentum aún está dado por $m\boldsymbol{v}$, pero si usamos la nueva m, ésta será:

$$\boldsymbol{p} = m\boldsymbol{v} = \frac{m_0\boldsymbol{v}}{\sqrt{1 - v^2/c^2}}. \tag{15.10}$$

Ésta es la modificación de Einstein a las leyes de Newton. En esta modificación, acción y reacción son todavía iguales (posiblemente puedan no ser en detalle,

pero sí en forma global), existirá la conservación del momentum en la misma forma anterior, pero la cantidad que se conserva no es el antiguo mv con masa constante, sino la cantidad indicada en (15.10) que contiene la masa modificada. Si se efectúa este cambio en la fórmula para el momentum, la conservación del momentum todavía funciona.

Veamos ahora cómo varía el momentum con la velocidad. En la mecánica newtoniana es proporcional a la velocidad y de acuerdo con (15.10) es casi el mismo en la mecánica relativista sobre un considerable intervalo de velocidades, pequeñas en comparación con c, dado que la raíz cuadrada difiere muy poco de 1. Pero si v es casi igual a c, la raíz cuadrada se aproxima a cero, y el momentum tiende por lo tanto a infinito.

¿Qué pasará si una fuerza constante actúa sobre un cuerpo durante un tiempo largo? En la mecánica newtoniana el cuerpo aumenta su velocidad hasta que camina más rápido que la luz. Pero esto es imposible en la mecánica relativista. En relatividad, el cuerpo no aumenta su velocidad, sino su momentum, que puede crecer continuamente porque su masa crece. Después de cierto tiempo, ya no existe prácticamente aceleración en el sentido de cambio de velocidad, pero el momentum continúa aumentando. Naturalmente si una fuerza produce un cambio muy pequeño en la velocidad de un cuerpo, nosotros decimos que el cuerpo tiene mucha inercia y esto es exactamente lo que dice nuestra fórmula para la masa relativista [vease ecuación (15.10)], dice que la inercia es muy grande cuando v es casi igual a c. Como ejemplo de este efecto puede mencionarse que para detectar electrones de alta velocidad en el sincrotrón usado aquí en el Caltech, necesitamos un campo magnético que es 2 000 veces más intenso que el valor que se esperaría con base de las leyes de Newton. En otras palabras, la masa de los electrones en el sincrotón es 2 000 veces mayor que su masa normal, ¡y es tan grande como la de un protón! Que m sea 2 000 veces m_0 significa que $1 - v^2/c^2$ debe ser $1/4\,000\,000$ y esto significa que v difiere de c por una parte en 8 000 000, es decir, la velocidad de los electrones se aproxima bastante a la luz. Si los electrones y la luz partieran ambos del sincrotón (estimado a 200 metros de distancia) y van al laboratorio Bridge, ¿quién llegaría primero? Naturalmente la luz, porque siempre la luz viaja más rápidamente.* ¿Cuánto antes? Esto es muy difícil de decir; en cambio, diremos en qué distancia se adelanta la luz: ¡es alrededor de 1/4 de décima de milímetro o 1/4 del espesor de una hoja de papel! Cuando los electrones se mueven así de rápido sus masas resultan enormes, pero nunca su velocidad puede exceder la velocidad de la luz.

Busquemos ahora otras consecuencias del cambio relativista de la masa. Consideremos el movimiento de las moléculas en un pequeño tanque de gas. Si se calienta el gas, aumenta la velocidad de las moléculas, y por eso también la masa crece y el gas es más pesado. Se puede obtener una fórmula aproximada para indicar el aumento de la masa para el caso de velocidades pequeñas, desarrollando $m_0/\sqrt{1 - v^2/c^2} = m0(1 - v^2/c^2)^{1/2}$ en una serie de potencias, usando el teorema del binomio. Obtendremos,

$$m_0(1 - v^2/c^2)^{-1/2} = m_0(1 + \tfrac{1}{2}v^2/c^2 + \tfrac{3}{8}v^4/c^4 + \cdots).$$

Vemos claramente de la fórmula que la serie converge rápidamente cuando v es pequeño, y los términos después de los dos primeros son despreciables. Así podemos escribir:

$$m \cong m_0 + \tfrac{1}{2}m_0v^2\left(\frac{1}{c^2}\right) \tag{15.11}$$

donde el segundo término del segundo miembro da el aumento de la masa debido a la velocidad molecular. Cuando la temperatura aumenta v^2 crece proporcionalmente, y así podemos decir que el aumento de la masa es proporcional al aumento de la temperatura. Pero dado que $\frac{1}{2}m_0v^2$ representa la energía cinética en el sentido anticuado newtoniano, podemos decir también que el aumento de

* En realidad los electrones ganarían la carrera contra la luz visible debido al índice de refracción del aire. Un rayo gamma lo haría mejor.

la masa de todo el gas es igual al aumento de la energía cinética dividido entre c^2, o bien $\Delta m = \Delta(EC)/c^2$.

15-9 Equivalencia de masa y energía

La observación de arriba condujo a Einstein a la sugerencia que la masa de un cuerpo se puede expresar de una manera más simple que por medio de la fórmula (15.1), diciendo que la masa es igual al contenido energético total dividido entre c^2. Si la ecuación (15.11) se multiplica por c^2 resulta:

$$mc^2 = m_0c^2 + \tfrac{1}{2}m_0v^2 + \cdots \tag{15.12}$$

Aquí el primer miembro da la energía total de un cuerpo, y en el último término reconocemos la energía cinética ordinaria. Einstein interpretó el término grande y constante m_0c^2, que forma parte de la energía total del cuerpo, como una energía intrínseca conocida como "energía de reposo".

Estudiemos más las consecuencias que resultan al suponer con Einstein que *la energía de un cuerpo es siempre* mc^2. Como un resultado interesante encontramos la fórmula (15.1) para la variación de la masa con la velocidad, que hasta ahora fue una mera suposición. Comenzamos considerando el cuerpo en reposo, cuando su energía es m_0c^2. Después aplicamos una fuerza al cuerpo, que le hace mover, dándole energía cinética, entonces, como la energía ha aumentado, también la masa ha aumentado —esto está implícito en la suposición original—. Mientras la fuerza continúa actuando, la energía y la masa continúan aumentando. Hemos visto ya (capítulo 13) que el cambio de energía con el tiempo es igual a la fuerza multiplicada por la velocidad, o

$$\frac{dE}{dt} = \boldsymbol{F}\cdot\boldsymbol{v}. \tag{15.13}$$

Además tenemos —capítulo 9, ecuación (9.1)— que $F = d(mv)/dt$. Cuando se juntan estas relaciones con la definición de E, la ecuación (15.13) da

$$\frac{d(mc^2)}{dt} = \boldsymbol{v}\cdot\frac{d(m\boldsymbol{v})}{dt}. \tag{15.14}$$

Queremos despejar m de esta ecuación. Para hacer esto usamos primero el truco matemático de multiplicar ambos miembros por $2m$, lo que cambia la ecuación

$$c^2(2m)\,\frac{dm}{dt} = 2m\boldsymbol{v}\cdot\frac{d(m\boldsymbol{v})}{dt}. \tag{15.15}$$

Tenemos que deshacernos de las derivadas, lo que puede lograrse integrando ambos miembros. La cantidad $(2m)dm/dt$ se puede reconocer como la derivada de m^2 respecto al tiempo. De esta manera la ecuación (15.15) es lo mismo que:

$$c^2\,\frac{d(m^2)}{dt} = \frac{d(m^2v^2)}{dt}. \tag{15.16}$$

Si las derivadas de dos cantidades son iguales, las cantidades mismas difieren a lo sumo en una constante, por ejemplo C. Esto nos permite escribir

$$m^2c^2 = m^2v^2 + C. \tag{15.17}$$

Es necesario definir más explícitamente la constante C. Dado que la ecuación (15.17) debe ser válida para todas las velocidades, podemos elegir el caso especial cuando $v = 0$, y decir que en este caso la masa es m_0. Sustituyendo estos valores en la ecuación (15.17) se obtiene:

$$m_0^2c^2 = 0 + C.$$

Ahora podemos usar este valor de C en la ecuación (15.17), lo que da

$$m^2c^2 = m^2v^2 + m_0^2c^2. \tag{15.18}$$

Dividiendo por c^2 y reordenando términos resulta:

$$m^2(1 - v^2/c^2) = m_0^2,$$

de lo cual obtenemos

$$m = m_0/\sqrt{1 - v^2/c^2}. \tag{15.19}$$

Ésta es la fórmula (15.1), y es exactamente lo necesario para la concordancia entre masa y energía en la ecuación (15.12).

Ordinariamente estos cambios de energía representan cambios extremadamente pequeños en la masa, porque en la mayoría de los casos no podemos generar mucha energía de una cierta cantidad de material; pero, por ejemplo, en una bomba atómica, de una energía explosiva equivalente a 20 kilotoneladas de TNT, se puede demostrar que el polvo después de la explosión es un gramo más liviano que la masa inicial del material en reacción, de acuerdo con la relación $\Delta E = \Delta(mc^2)$. Esta teoría de equivalencia de masa y energía ha sido verificada maravillosamente con experimentos en los cuales se aniquila la materia —convirtiéndola totalmente en energía, un electrón y un positrón llegan al reposo, cada uno con una masa de reposo m_0—. Cuando se juntan se desintegran y emergen dos rayos gamma cada uno con una energía m_0c^2. Este experimento proporciona una determinación directa de la energía asociada a la masa en reposo de una partícula.

16

Energía relativista y momentum

16-1 La relatividad y los filósofos

En este capítulo continuaremos discutiendo el principio de relatividad de Einstein y Poincaré, y la manera como afecta nuestras ideas de la física y otras ramas del pensamiento humano.

Poincaré hizo la siguiente exposición del principio de relatividad: "De acuerdo con el principio de relatividad, las leyes de los fenómenos físicos deben ser las mismas para un observador fijo que para un observador que tiene un movimiento uniforme de traslación relativo a él, de manera que no tengamos, ni siquiera posiblemente, ninguna manera de discernir si nosotros somos llevados o no en este movimiento".

Cuando esta idea se propagó por el mundo, causó un gran revuelo entre los filósofos, especialmente "los filósofos de salón" quienes dijeron: "Ah, es muy simple, la teoría de Einstein dice que ¡todo es relativo!" Realmente una sorprendente cantina de filósofos, no solamente aquellos que se encuentran en fiestas (pero con el ánimo de no avergonzarlos, sencillamente los llamaremos "filósofos de salón"), dirán "que todo es relativo, es una consecuencia de Einstein y tiene una profunda influencia en nuestras ideas". Además dicen: "Se ha demostrado en física que los fenómenos dependen del sistema de referencia". Se escucha esto muy a menudo, pero es difícil saber lo que significa. Probablemente los sistemas de referencia a que se refirieron originalmente eran los sistemas de coordenadas que usamos en el análisis de la teoría de la relatividad. De manera de que el hecho que "las cosas dependen de su sistema de referencia" ha tenido una profunda influencia en el pensamiento moderno. Uno podría muy bien preguntarse por qué, ya que, después de todo, que las cosas dependen del punto de vista de uno es una idea tan simple, que ciertamente no puede haber sido necesario todo el trastorno de la teoría de la relatividad física, para descubrirla. Que las cosas que uno ve dependen de su sistema de referencia es ciertamente conocido por cualquier persona que camine, ya que ve un peatón que se acerca primero por el frente y después por detrás, no hay nada más profundo en casi toda la filosofía que se dice que ha venido de la teoría de relatividad, que la observación de que "una persona luce diferente desde el frente que desde la parte de atrás". La vieja historia sobre el elefante descrito en forma diferente por varios ciegos es otro ejemplo, quizá de la teoría de la relatividad desde el punto de vista de los filósofos.

Pero realmente debe haber cosas más profundas en la teoría de la relatividad que la sencilla observación que "una persona luce diferente desde el frente que desde la parte de atrás". Por supuesto que la relatividad es más profunda *ya que podemos hacer predicciones definidas con ella.* Sería realmente notable si pudiéramos predecir el comportamiento de la naturaleza de una observación tan simple solamente.

Hay otras escuelas de filósofos que se sienten incómodos con la teoría de relatividad que asevera que no podemos determinar nuestra velocidad absoluta sin mirar a algo afuera, y los que dirían: "Es evidente que uno no puede medir su velocidad sin mirar afuera. Es patente que *no tiene sentido* hablar de la velocidad de algo sin mirar afuera, los físicos son bastante estúpidos al pensar diferente, pero ya han empezado a vislumbrar que así debe ser. Si nosotros los filósofos tan sólo nos hubiéramos dado cuenta de cuáles eran los problemas que tenían los físicos, podríamos haber decidido inmediatamente y mediante trabajo mental que es imposible decir con qué rapidez uno se está moviendo sin mirar afuera y podríamos haber hecho una enorme contribución a la física". Estos filósofos

están siempre con nosotros, luchando en la periferia para tratar de decirnos algo, pero nunca realmente entienden las sutilezas y profundidades del tema.

Nuestra incapacidad de detectar movimiento absoluto es el resultado de *los experimentos* y no el resultado de raciocinio solamente, como se puede ilustrar fácilmente. En primer lugar, Newton creyó que uno no puede decir a qué velocidad va si uno se está moviendo con velocidad uniforme en una línea recta. Realmente, Newton fue quien primero estableció el principio de relatividad y una cita hecha en el capítulo precedente era una observación de Newton. ¿Por qué, entonces, los filósofos no hicieron todo este alboroto porque "todo es relativo" o algo por el estilo, en tiempo de Newton? Porque no fue hasta que se desarrolló la teoría de Maxwell, que hubo leyes físicas que sugirieron que uno *podría* medir su velocidad sin mirar afuera, pronto se encontró *experimentalmente* que uno *no* podía.

Ahora bien, ¿*es* absolutamente, definitivamente, filosóficamente *necesario* que uno no sea capaz de decir con qué velocidad se está moviendo sin mirar hacia afuera? Una de las consecuencias de la relatividad fue el desarrollo de una filosofía que decía: "¡Usted puede definir sólo lo que puede medir! Ya que es patente que uno no puede medir una velocidad sin ver respecto a qué la está midiendo, está claro que no tiene *sentido* la velocidad absoluta. Los físicos deberían haberse dado cuenta que pueden hablar solamente de aquello que pueden medir". Pero *ahí está el problema*, si uno *puede definir* o no velocidad absoluta es lo mismo que el problema de si uno puede o no *detectar en un experimento*, sin mirar afuera, si uno se está moviendo. En otras palabras, si una cosa es medible o no, no es algo que se decide *a priori* por el solo pensamiento, sino algo que puede ser decidido solamente por el experimento. Dado el hecho de que la velocidad de la luz es 300 000 km/s, uno va a encontrar pocos filósofos que vayan a decir calmadamente que es patente que si la luz va a 300 000 km/s dentro de un auto y el auto va a 200 000 km/s que la luz también va a 300 000 km/s con respecto a un observador en el suelo. Ésto es un hecho chocante para ellos, los mismos que claman que es evidente, encuentran que no lo es cuando se les da un hecho específico.

Finalmente, hay una filosofía que dice que uno no puede detectar *ningún* movimiento excepto mirando hacia afuera. Simplemente, esto no es verdadero en física. Cierto, uno no puede percibir movimiento *uniforme* en una *línea recta*, pero si toda la pieza estuviera *rotando* lo sabríamos con toda seguridad, ya que todo el mundo sería arrojado hacia la pared —habría toda clase de efectos "centrífugos"—. Que la Tierra gira sobre su eje puede determinarse sin mirar a las estrellas, mediante el así llamado péndulo de Foucault, por ejemplo. Por lo tanto, no es cierto que "todo es relativo", es solamente la *velocidad uniforme* la que no se puede detectar sin mirar hacia afuera. La *rotación* uniforme alrededor de un eje se *puede* detectar. Cuando se le dice esto a un filósofo queda muy contrariado porque realmente no lo entendió, porque para él parece imposible que uno sea capaz de determinar la rotación alrededor de un eje sin mirar hacia afuera. Si el filósofo es suficientemente bueno, después de algún tiempo puede volver y decir: "Yo entiendo. Realmente no tenemos una rotación absoluta, estamos realmente rotando *relativo a las estrellas*, vea usted, y alguna influencia ejercida por las estrellas sobre el objeto debe causar la fuerza centífruga".

Ahora bien, que nosotros sepamos eso es cierto, no tenemos ninguna manera, en estos momentos, de determinar si habría habido fuerza centrífuga, si no hubiera estrellas y nebulosas alrededor. No hemos podido hacer la experiencia de sacar todas las nebulosas y después medir nuestra rotación, así que sencillamente no sabemos. Debemos admitir que el filósofo puede tener razón. Vuelve entonces deleitado y dice: "Es absolutamente necesario que el mundo resulte así en última instancia: rotación *absoluta* no significa nada, es solamente *relativa* a las nebulosas". Entonces le decimos, "*ahora*, mi amigo, ¿es o no evidente que una velocidad uniforme en línea recta, *relativo a las nebulosas*, no debería producir efectos en el auto?" Ahora que el movimiento ya no es absoluto, sino que es un movimiento *relativo a las nebulosas*, la pregunta se hace misteriosa y una pregunta que puede ser contestada solamente mediante experimentos.

¿Cuáles *son* entonces las influencias filosóficas de la teoría de la relatividad? Si nos limitamos a influencias en el sentido de *qué clase de nuevas ideas y sugerencias*

hace al físico el principio de relatividad, podríamos describir algunas como sigue. El primer descubrimiento es, esencialmente, que aun aquellas ideas que se han mantenido por mucho tiempo y que han sido verificadas precisamente, pueden estar equivocadas. Fue un descubrimiento chocante, por supuesto, que las leyes de Newton están equivocadas, después de tantos años que parecían precisas. Por supuesto, que los experimentos no estaban mal, sólo que fueron hechos en un intervalo limitado de velocidades, tan pequeñas que los efectos relativísticos no podían evidenciarse. Sin embargo, ahora tenemos un punto de vista mucho más humilde de nuestras leyes físicas: ¡todo *puede* estar mal!

En segundo lugar, si tenemos un conjunto de ideas "extrañas", como que el tiempo avanza más despacio cuando uno se mueve, etc., bien que nos *gusten o no*, es una pregunta fuera de lugar. La única pregunta pertinente es si las ideas son compatibles con lo que se encuentra experimentalmente. En otras palabras, las "ideas extrañas" necesitan solamente estar de acuerdo con los *experimentos*, y la única razón que tenemos para discutir el comportamiento de los relojes y lo demás es demostrar que, aunque la noción de la dilatación del tiempo es extraña, es *consistente* con la manera en que medimos el tiempo.

Finalmente, hay una tercera sugerencia que es un poco más técnica, pero que ha resultado ser de enorme utilidad en nuestro estudio de otras leyes físicas, y ella es *observar la simetría de las leyes* o, más específicamente, buscar los procedimientos de transformar las leyes dejándoles la misma forma. Cuando discutimos la teoría de los vectores, notamos que las leyes fundamentales del movimiento no cambian cuando rotamos el sistema de coordenadas, y ahora encontramos que no cambian cuando cambiamos las variables de espacio y tiempo de una manera particular, dada por la transformación de Lorentz. Por lo tanto, esta idea, de estudiar los esquemas u operaciones bajo los cuales las leyes fundamentales no cambian, ha demostrado ser muy útil.

16-2 La paradoja de los mellizos

Para continuar nuestra discusión de la transformación de Lorentz y efectos relativísticos, consideremos la famosa "paradoja" de Pedro y Pablo, que se supone son mellizos, nacidos al mismo tiempo. Cuando tienen la edad suficiente para manejar una nave espacial, Pablo hace un viaje a gran velocidad. Ya que Pedro, que queda en tierra, ve a Pablo viajar a tan gran velocidad, todos los relojes de Pablo parece que se atrasan, sus latidos son más lentos, sus pensamientos van más despacio, todo va más lento desde el punto de vista de Pedro. ¡Por supuesto, que Pablo no nota nada fuera de lo común, pero si viaja de un lado a otro por un tiempo y después vuelve, va a ser más joven que Pedro, el hombre que se quedó en la Tierra! Esto es realmente verdadero, es una de las consecuencias de la teoría de la relatividad que ha sido demostrada claramente. Así como los muones duran más cuando se están moviendo, también Pablo va a durar más mientras se mueve. A esto llaman "paradoja" solamente aquella gente que cree que el principio de relatividad significa que *todo movimiento* es relativo, ellos dicen, "eh eh eh, desde el punto de vista de Pablo, ¿no podríamos decir que es *Pedro* el que se está moviendo y, por lo tanto, no debería parecer que él envejece más lentamente? Por simetría, el único resultado posible es que tengan la misma edad cuando se encuentren". Pero para que se junten y se pueda hacer la comparación, Pablo debe o detenerse al final del viaje y hacer una comparación de relojes, o más sencillo, volver y el que vuelve tiene que ser el hombre que se estaba moviendo, y esto él lo sabe, porque tuvo que cambiar el sentido de su movimiento. Cuando cambió el sentido, todo tipo de cosas poco comunes sucedieron en su nave espacial —los cohetes se apagaron, las cosas se apretaron contra una pared, etc.— mientras que Pedro no sintió nada.

Por lo tanto, la manera de exponer la regla es decir que *el hombre que sintió las aceleraciones*, el que vio las cosas caer sobre las paredes, etc., es el que va a ser menor, ésa es la diferencia entre ellos en un sentido "absoluto", y es, ciertamente, correcto. Cuando discutimos el hecho que los muones que se mueven viven más, usamos como ejemplo su movimiento rectilíneo en la atmósfera. Pero

también podemos hacer muones en un laboratorio y hacerlos seguir una curva mediante un imán, y aun con este movimiento acelerado, duran exactamente lo mismo que cuando se estaban moviendo en línea recta. Aunque nadie ha hecho un experimento explícitamente de manera que pudiéramos deshacernos de la paradoja, uno podría comparar un muon que se ha dejado quieto con uno que ha dado una vuelta completa en un círculo, y seguramente descubriría que el que se movió en círculo duró más. Aunque realmente no hemos realizado un experimento usando un círculo completo, esto no es necesario, por supuesto, pues todo ajusta perfectamente. Esto puede no satisfacer a aquellos que insisten en que cada hecho sea demostrado directamente, pero nosotros predecimos con toda confianza el resultado del experimento en el cual Pablo se mueve en un círculo completo.

16-3 Transformación de de velocidades

La mayor diferencia entre la relatividad de Einstein y la relatividad de Newton es que las leyes de transformación, que conectan las coordenadas y tiempos entre sistemas con movimiento relativo, son diferentes. La ley de transformación correcta, la de Lorentz, es:

$$\begin{aligned} x' &= \frac{x - ut}{\sqrt{1 - u^2/c^2}}, \\ y' &= y, \\ z' &= z, \\ t' &= \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}}. \end{aligned} \tag{16.1}$$

Estas ecuaciones corresponden al caso relativamente simple en que el movimiento relativo de los observadores se realiza a lo largo de su eje común x. Por supuesto otras direcciones de movimiento son posibles, pero la transformación de Lorentz más general es bastante complicada, con las cuatro cantidades mezcladas entre sí. Continuaremos usando esta manera más simple, ya que contiene los aspectos esenciales de la relatividad.

Discutamos más sobre las consecuencias de esta transformación. Primero, es interesante resolver estas ecuaciones a la inversa. Esto es, aquí tenemos un conjunto de ecuaciones lineales, cuatro ecuaciones con cuatro incógnitas y se las puede invertir para despejar x, y, z, t en función de x', y', z', t'. El resultado es muy interesante, ya que nos dice cómo se ve un sistema de coordenadas "en reposo" desde el punto de vista de uno que se está "moviendo". Por supuesto, ya que los movimientos son relativos y de velocidad uniforme, el hombre que se "muev" puede decir, si desea, que realmente es la otra persona la que se mueve, mientras que él está en reposo y ya que se está moviendo en dirección opuesta, él debe obtener la misma transformación, pero con signo contrario de la velocidad. Esto es precisamente lo que encontramos al efectuar los cálculos, de manera que esto es compatible. ¡Si no hubiera resultado así, ahí sí que habíamos tenido una causa real para preocuparnos!

$$\begin{aligned} x &= \frac{x' + ut'}{\sqrt{1 - u^2/c^2}}, \\ y &= y', \\ z &= z', \\ t &= \frac{t' + ux'/c^2}{\sqrt{1 - u^2/c^2}}. \end{aligned} \tag{16.2}$$

A continuación discutimos el interesante problema de la suma de velocidades en relatividad. Recordemos que el enigma original era que la luz viaja a 300 000 km/s en todos los sistemas aunque está en movimiento relativo. Esto es un caso particular del problema más general ilustrado por lo que sigue. Supongamos

que un objeto va a 200 000 km/s dentro de una nave espacial y la nave espacial a su vez a 200 000 km/s. ¿Con qué velocidad se mueve el objeto dentro de la nave espacial desde el punto de vista de un observador que está afuera? Nos gustaría decir 400 000 km/s, que es mayor que la velocidad de la luz. Esto es muy inquietante, ¡porque se supone que no puede ir a mayor velocidad que la luz! El problema general es el siguiente.

Supongamos que el objeto dentro de la nave, desde el punto de vista del hombre dentro de ella, se mueve con velocidad v, y que la nave misma tenga una velocidad u respecto a la Tierra. Queremos saber con qué velocidad v_x este objeto se está moviendo desde el punto de vista del hombre en la Tierra. Esto es, por supuesto, un caso especial en el cual la dirección es a lo largo de la dirección x. También va a haber una transformación para la velocidad en la dirección y, o bien, en cualquier ángulo. Éstas se pueden desarrollar a medida que se necesite. Dentro de la nave espacial la velocidad es $v_{x'}$, lo que significa que el desplazamiento x' es igual a la velocidad por el tiempo:

$$x' = v_{x'}t'. \tag{16.3}$$

Ahora nos queda solamente calcular cuáles son la posición y el tiempo desde el punto de vista de un observador ubicado afuera para un objeto que tiene la relación (16.2) entre x' y t. Así que simplemente sustituimos (16.3) en (16.2) y obtenemos:

$$x = \frac{v_{x'}t' + ut'}{\sqrt{1 - u^2/c^2}}. \tag{16.4}$$

Pero aquí encontramos x expresado en función de t'. Para obtener la velocidad que detecta al hombre de afuera, ¡debemos dividir *su distancia* entre *su tiempo*, y no entre el tiempo del otro hombre! Por lo tanto, debemos también calcular el tiempo como visto desde afuera, que es:

$$t = \frac{t' + u(v_{x'}t')/c^2}{\sqrt{1 - u^2/c^2}}. \tag{16.5}$$

Ahora debemos encontrar el cociente entre x y t que es:

$$v_x = \frac{x}{t} = \frac{u + v_{x'}}{1 + uv_{x'}/c^2}, \tag{16.6}$$

habiendo simplificado las raíces cuadradas. Ésta es la ley que buscamos, la velocidad resultante, la "suma" de dos velocidades, no es simplemente la suma algebraica de dos velocidades (sabemos que no puede ser o estaríamos en aprietos), sino que está "corregida" por $1 + uv/c^2$.

Veamos ahora lo que pasa. Supongan que se mueven dentro de la nave espacial a la mitad de la velocidad de la luz, y que la nave espacial misma va a la mitad de la velocidad de la luz. Luego u es $\frac{1}{2}c$ y v es $\frac{1}{2}c$, pero en el denominador uv es un cuarto, de manera que,

$$v = \frac{\frac{1}{2}c + \frac{1}{2}c}{1 + \frac{1}{4}} = \frac{4c}{5}.$$

Así que, en relatividad, "un medio" más "un medio" no es "un entero", es sólo "4/5". Por supuesto que velocidades bajas pueden sumarse con toda facilidad de la manera acostumbrada, porque mientras las velocidades sean pequeñas comparadas con la velocidad de la luz, nos podemos olvidar del factor $(1 + uv/c^2)$, pero las cosas son bastante diferentes e interesantes a alta velocidad.

Tomemos un caso límite. Para entretenernos, supongamos que el hombre estuviera observando la *luz* dentro de la nave espacial. En otras palabras $v = c$, y, sin embargo, la nave espacial se está moviendo. ¿Qué le va a parecer al hombre en la Tierra? La respuesta va a ser:

$$v = \frac{u + c}{1 + uc/c^2} = c\,\frac{u + c}{u + c} = c.$$

Por lo tanto, si algo se está moviendo a la velocidad de la luz dentro de la nave, ¡va a parecer estarse moviendo también con la velocidad de la luz desde el punto de vista del hombre en la Tierra! Esto está bueno, porque es en realidad para lo que la teoría de relatividad de Einstein estaba diseñada en primer lugar, ¡así que *más valía* que resultara!.

Por supuesto, hay casos en los cuales el movimiento no es en la dirección de la traslación uniforme. Por ejemplo, podría haber un objeto dentro de la nave que se está moviendo "hacia arriba" con velocidad $v_{y'}$ respecto a la nave, y la nave se está moviendo "horizontalmente". Ahora bien, hacemos el mismo desarrollo, sólo usando y en vez de x, con resultado

$$y = y' = v_{y'}t',$$

de manera que $v_{x'} = 0$,

$$v_y = \frac{y}{t} = v_{y'}\sqrt{1 - u^2/c^2}. \tag{16.7}$$

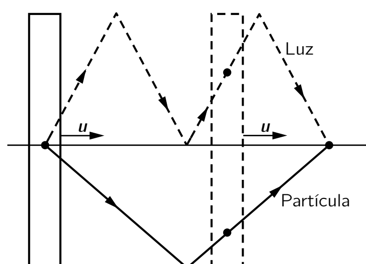


Figura 16-1. Trayectorias descritas por un rayo de luz y una partícula dentro de un reloj en movimiento.

Luego, una velocidad transversal ya no es v_y, sino $v_y\sqrt{1 - u^2/c^2}$. Encontramos este resultado sustituyendo y combinando las ecuaciones de transformación, pero también podemos ver el resultado directamente con el principio de relatividad por la siguiente razón (es siempre conveniente pensar de nuevo para ver si podemos encontrar la razón). Ya hemos visto (figura 15-3) cómo un posible reloj podría trabajar cuando se está moviendo, la luz parece desplazarse oblicuamente con velocidad c en el sistema fijo, cuando simplemente se desplaza en forma vertical en el sistema móvil. Encontramos que la *componente vertical* de la velocidad en el sistema fijo es menor que la de la luz por un factor $\sqrt{1 - u^2/c^2}$ [véase ecuación (15.3)]. Pero supongan ahora que dejamos una partícula material ir hacia atrás y hacia adelante en este mismo "reloj", pero a una fracción entera $1/n$ de la velocidad de la luz (figura 16-1). Luego, cuando la partícula ha ido hacia atrás y hacia adelante una vez, la luz va a ir exactamente n veces. Es decir, cada "tic" del reloj de partícula va a coincidir con el n-ésimo "tic" del reloj de luz. *Este hecho debe seguir siendo cierto cuando todo el sistema se está moviendo*, porque el fenómeno físico de coincidencia va a ser una coincidencia en cualquier sistema. Por lo tanto, ya que la velocidad c_y es menor que la velocidad de la luz, ¡la velocidad v_y de la partícula debe ser menor que la velocidad correspondiente en la misma razón de la raíz cuadrada! Por eso es que la raíz cuadrada está en cualquier velocidad vertical.

16-4 Masa relativista

Aprendimos en el último capítulo que la masa de un objeto aumenta con la velocidad, pero no se dio ninguna demostración en el sentido que no dimos razonamientos análogos a aquellos acerca de la manera en que los relojes deben comportarse. Sin embargo, *podemos* demostrar que, como consecuencia de la relatividad más algunas otras hipótesis razonables, la masa debe variar de esta manera. (Tenemos que decir "algunas otras hipótesis" porque no podemos probar nada a menos que tengamos algunas leyes que suponemos que sean ciertas, si queremos hacer deducciones con algún sentido.) Para evitar tener que estudiar las leyes de transformación de la fuerza, vamos a analizar una *colisión*, donde no necesitamos saber nada sobre las leyes de la fuerza, solamente vamos a suponer la conservación del momentum y de la energía. También vamos a suponer que el momentum de una partícula que se mueve es un vector y está siempre dirigido en la dirección de la velocidad. Sin embargo, no vamos a suponer que momentum es una *constante* multiplicada por la velocidad, como hizo Newton, sino que es alguna *función* de la velocidad. Escribimos, pues, el vector momentum como un cierto coeficiente multiplicado por el vector velocidad:

$$\boldsymbol{p} = m_v\boldsymbol{v}. \tag{16.8}$$

Pusimos un subíndice v en el coeficiente para recordarnos que es una función de la velocidad, y vamos a convenir llamar "masa" a este coeficiente m. Por

supuesto, cuando la velocidad es pequeña, es la misma masa que mediríamos en los experimentos de movimientos lentos a que estamos acostumbrados. Ahora vamos a tratar de demostrar que la fórmula para mv debe ser $m_0/\sqrt{1-v^2/c^2}$, argumentando a partir del principio de la relatividad que las leyes de la física deben ser las mismas en todo sistema de coordenadas.

Supongamos que tenemos dos partículas, por ejemplo dos protones, que son absolutamente iguales y que se están acercando con velocidades exactamente iguales. Su momentum total es cero. Ahora bien, ¿qué puede suceder? Después de la colisión, sus direcciones de movimiento deben ser exactamente opuestas porque si no fueran exactamente opuestas, habría un vector momentum total diferente de cero, y el momentum no se conservaría. Deben tener también la misma velocidad, ya que son objetos exactamente similares, de hecho, deben tener la misma velocidad con que partieron ya que suponemos que la energía se conserva en estas colisiones. Por lo tanto, el diagrama de una colisión elástica, una colisión reversible, va a ser algo como en la figura 16-2(a): todas las flechas son del mismo largo, todas las velocidades son iguales. Vamos a suponer que podemos arreglar las colisiones de manera tal que cualquier ángulo θ puede darse, y cualquier velocidad puede usarse en ellas. Enseguida, vemos que este mismo impacto puede ser visualizado en forma diferente rotando los ejes, y por conveniencia los *vamos* a rotar de manera que la horizontal lo separe en forma pareja, como en la figura 16-2(b). Es la misma colisión dibujada de nuevo, solamente que los ejes se han rotado.

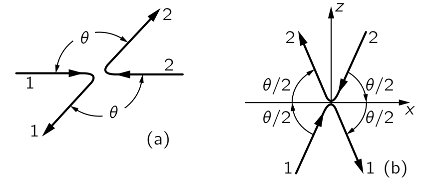


Figura 16-2. Dos vistas de una colisión elástica entre dos objetos iguales moviéndose con la misma velocidad en sentidos contrarios.

Ahora viene el truco, miremos esta colisión desde el punto de vista de alguien que viaja en un auto que se mueve con una velocidad igual a la componente horizontal de la velocidad de una partícula. ¿Cómo aparece la colisión? Parece como si la partícula 1 fuera derecho hacia arriba, ya que ha perdido su componente horizontal y volviera derecho hacia deabajo de nuevo porque tampoco tiene esa componente. O sea que la colisión aparece como se muestra en la figura 16-3(a). La partícula 2, sin embargo, iba en el otro sentido, y a medida que pasamos cerca de ella, parece que tuviera una tremenda velocidad y un ángulo menor, pero podemos apreciar que los ángulos antes y después del impacto son los *mismos*. Designemos con u la componente horizontal de la velocidad de la partícula 2, y con w la velocidad vertical de la partícula 1.

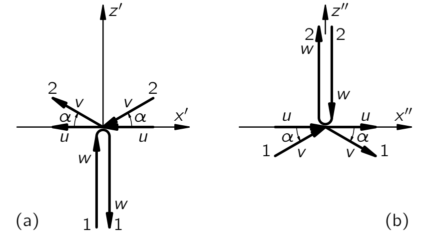


Figura 16-3. Dos vistas más de la colisión, desde autos en movimiento.

La pregunta ahora es, ¿cuál es la velocidad vertical $u\tan\alpha$? Si supiéramos lo anterior, obtendríamos la expresión correcta del momentum, usando la ley de conservación del momentum en la dirección vertical. Evidentemente la componente horizontal del momentum se conserva, es la misma antes y después de la colisión para ambas partículas, y es cero para la partícula 1. Por lo tanto, necesitamos usar la ley de conservación sólo para la velocidad hacia arriba $u\tan\alpha$. ¡Pero *podemos* obtener la velocidad hacia arriba simplemente observando la misma colisión yendo en el otro sentido! Si miramos la colisión de la figura 16-3(a) desde un auto moviendose a la izquierda con velocidad u, vemos la misma colisión, con la diferencia que está "dada vuelta", como se muestra en la figura 16-3(b). Ahora la partícula 2 es la que va de arriba a abajo con velocidad w, la partícula 1 ha tomado la velocidad horizontal u. Por supuesto, ahora *sabemos* lo que es la velocidad $u\,\mathrm{tg}\,\alpha : \sqrt{1-u^2/c^2}$ [véase ecuación (16.7)]. Sabemos que el cambio del momemum vertical de la partícula que se mueve verticalmente es:

$$\Delta p = 2m_w w$$

(2 es porque se mueve hacia arriba y después hacia abajo). La partícula que se mueve en forma oblicua tiene una cierta velocidad v cuyas componentes hemos encontrado que son u y $w\sqrt{1-u^2/c^2}$, y cuya masa es m_v. El cambio del momentum *vertical* de esta partícula es por lo tanto $\Delta p' = 2m_v w\sqrt{1-u^2/c^2}$ porque, de acuerdo con la ley que hemos supuesto (16.8), la componente del momentum es siempre la masa correspondiente a la magnitud de la velocidad multiplicada por la componente de la velocidad en la dirección de interés. Por lo tanto, para que el momentum total sea cero, los momentum verticales deben anularse y el cociente entre la masa que se mueve con velocidad v y la masa que se mueve con velocidad w debe ser entonces

$$\frac{m_w}{m_v} = \sqrt{1 - u^2/c^2}. \tag{16.9}$$

Tomemos el caso límite cuando w sea infinitesimal. Si w es muy pequeña, es claro que v y u son prácticamente iguales. En este caso, $m_w \to m_0$ y $m_v \to m_u$. El gran resultado es:

$$m_u = \frac{m_0}{\sqrt{1 - u^2/c^2}}. \tag{16.10}$$

Es un ejercicio interesante verificar si la ecuación (16.9) es realmente válida para valores arbitrarios de w, suponiendo que la ecuación (16.10) es la fórmula correcta para la masa. Nótese que la velocidad v que se necesita en la ecuación (16.9) puede calcularse del triángulo rectángulo:

$$v^2 = u^2 + w^2(1 - u^2/c^2).$$

Se encontrará que se verifica automáticamente, aunque lo usamos solamente en el caso límite de w pequeño.

Ahora aceptemos que el momentum se conserva y que la masa depende de la velocidad de acuerdo con (16.10) y sigamos observando qué más podemos concluir. Consideremos lo que se llama comúnmente una *colisión inelástica*. Para simplificar, vamos a suponer que dos objetos del mismo tipo, que se mueven en direcciones opuestas con la misma velocidad w, chocan y quedan pegados, transformándose en algún objeto nuevo, en reposo como se muestra en la figura 16-4(a). La masa m de cada uno corresponde a w, que según sabemos vale $m_0\sqrt{1 - w^2/c^2}$. Si suponemos la conservación del momentum y el principio de relatividad, podemos demostrar un hecho interesante acerca de la masa del nuevo objeto que acaba de formarse. Imaginemos una velocidad infinitesimal u formando un ángulo recto con w (podemos hacer lo mismo con valores finitos de u, pero es más fácil entender con una velocidad infinitesimal) y después miramos esta misma colisión mientras vamos en un ascensor con velocidad $-u$. Lo que vemos se muestra en la figura 16-4(b). El objeto compuesto tiene una masa desconocida M. Ahora el objeto 1 se mueve hacia arriba con una componente vertical de velocidad u y una componente horizontal que es prácticamente igual a w, y lo mismo hace el objeto 2. Después del impacto tenemos la masa M moviéndose hacia arriba con velocidad u, considerada muy pequeña comparada con la velocidad de la luz y pequeña también comparada con w. El momentum debe conservarse, así que hagamos una estimación del momentum en la dirección hacia arriba antes y después de la colisión. Antes de la colisión tenemos $p \approx 2m_w u$, y después de la colisión, el momentum es evidentemente $p' = M_u u$, pero M_u es esencialmente lo mismo que M_0 porque u es muy pequeña. Estos momenta deben ser iguales debido a la conservación del momentum, y por lo tanto

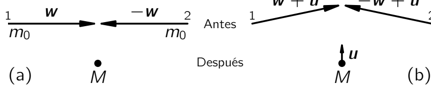


Figura 16-4. Dos vistas de una colisión inelástica entre dos objetos de igual masa.

$$M_0 = 2m_w. \tag{16.11}$$

La masa del objeto que se forma cuando dos objetos iguales chocan debe ser el doble de la masa de los objetos que se juntan. Ustedes podrían decir: "Sí, por supuesto, ésa es la conservación de la masa". Pero no "sí, por supuesto", tan fácilmente, porque *estas masas han sido aumentadas* respecto a las masas que tendrían si estuvieran en reposo y, sin embargo, ellas contribuyen al M total no sólo la masa que tienen cuando están en reposo, sino que *más*. Por muy sorprendente que esto pueda aparecer, para que la conservación del momentum resulte cuando los dos objetos se juntan, ¡la masa que forman debe ser mayor que la masa en reposo de los objetos, aunque los objetos estén en reposo después del choque!

16-5 Energía relativista

En el último capítulo demostramos que como consecuencia de la dependencia de la masa en la velocidad y de las leyes de Newton, los cambios en la energía

cinética de un objeto como resultado del trabajo total hecho por las fuerzas que actúan sobre él resultan ser:

$$\Delta T = (m_u - m_0)c^2 = \frac{m_0c^2}{\sqrt{1 - u^2/c^2}} - m_0c^2. \tag{16.12}$$

Fuimos aún más allá, y supusimos que la energía total es la masa total multiplicada por c^2. Ahora continuamos esta discusión.

Supongamos que nuestros dos objetos de masas iguales que chocan, todavía pueden ser "vistos" dentro de M. Por ejemplo, un protón y neutrón "se mantienen unidos", pero se están moviendo dentro de M. Por eso, aunque podríamos al principio esperar que la masa de M fuera $2m_0$, hemos encontrado que no es $2m_0$, sino $2m_w$. Ya que $2m_w$ es lo que hay adentro, pero $2m_0$ son las masas en reposo de lo que hay adentro, el *exceso* de masa del objeto compuesto es igual a la energía cinética que se llevó adentro. Esto significa, por supuesto, que la *energía tiene inercia*. En el último capítulo discutimos el calentamiento de un gas y mostramos eso porque las moléculas del gas que se están moviendo y las cosas que se mueven son más pesadas, cuando entregamos energía al gas sus moléculas se mueven más rápido y el gas se hace más pesado. Pero en realidad, el razonamiento es completamente general, y nuestra discusión del choque inelástico muestra que la masa está ahí, sea o no energía *cinética*. En otras palabras, si dos partículas se juntan y producen energía potencial o cualquier otra forma de energía, si los pedazos se frenan al subir planos inclinados, al realizar trabajo en contra de fuerzas internas, o lo que sea, entonces es todavía cierto que la masa es la energía total que se ha entregado. Vemos así que la conservación de la masa que dedujimos más arriba es equivalente a la conservación de la energía y, por lo tanto, no hay lugar en la teoría de la relatividad para choques completamente inelásticos como había en la mecánica newtoniana. De acuerdo con la mecánica newtoniana, está permitido que dos cosas choquen y así formen un objeto de masa $2m_0$ que no se distingue en nada del que resultaría al juntarlas despacio. Por supuesto, nosotros sabemos por la ley de conservación de la energía, que hay más energía cinética adentro pero que eso no afecta la masa, de acuerdo con las leyes de Newton. Pero ahora vemos que esto es imposible, debido a la energía cinética que interviene en la colisión, el objeto resultante va a ser más *pesado*, por lo tanto será un objeto *diferente*. Cuando juntamos estos objetos con cuidado forman algo cuya masa es $2m_0$, cuando los juntamos con fuerza forman algo cuya masa es mayor. Cuando la masa es diferente, podemos *darnos cuenta* de que es diferente. Por lo tanto, necesariamente la conservación de la energía debe cumplirse conjuntamente con la conservación del momentum en la teoría de la relatividad.

Esto tiene consecuencias interesantes. Por ejemplo, supongamos que tenemos un objeto cuya masa M se mide, y supongamos que algo sucede de forma que se rompe en dos pedazos iguales que se mueven con velocidad w, de manera que cada uno tiene una masa M_w. Supongamos ahora que estos pedazos tropiezan con suficiente material para frenarlos hasta detenerlos, entonces tendrán una masa m_0. ¿Cuánta energía entregarían al material una vez que se han detenido? Cada uno va a dar una cantidad $(m_w - m_0)c^2$, según el teorema que demostramos antes. Toda esta energía queda en el material de alguna manera, como calor, energía potencial, o lo que sea. Ahora bien, $2m_w = M$, de manera que la energía liberada es $E = (M - 2m_0)c^2$. Esta ecuación fue usada para estimar cuánta energía sería liberada por fisión en la bomba atómica, por ejemplo. (Aunque los fragmentos no son exactamente iguales, son casi iguales.) La masa del átomo de uranio era conocida —se había medido con anterioridad— y los átomos en los cuales se rompe, yodo, xenón, etc., eran todos de masa conocida. Por masa, no entendemos las masas mientras los átomos se están moviendo, sino las masas cuando los átomos están *en reposo*. En otras palabras, M y m_0 son conocidos. De manera que restando los dos números uno puede calcular cuánta energía será liberada si M se puede romper por la "mitad". Por esta razón, el pobre viejo Einstein fue llamado el "padre" de la bomba atómica en todos los periódicos.

Por supuesto, todo lo que eso significa era que él podía decirnos con anticipación cuánta energía se podría liberar si le dijéramos qué proceso iba a ocurrir. La energía que debería ser liberada cuando un átomo de uranio se fisionara se estimó seis meses antes de la primera prueba directa, y tan pronto como la energía fue liberada, alguien la midió directamente (y si la fórmula de Einstein no hubiera dado resultado, ellos la habrían medido de todos modos), y en el momento que la midieron ya no necesitaron más la fórmula. Lógicamente, no debemos quitar mérito a Einstein, sino criticar los periódicos y muchas descripciones populares de qué cosa produce qué cosa en la historia de la física y de la tecnología. El problema de cómo obtener que algo suceda de una manera efectiva y rápida es algo totalmente diferente.

El resultado es igualmente significativo en química. Por ejemplo, si pesáramos la molécula de dióxido de carbono y comparáramos su masa con la del carbono y del oxígeno, sería posible averiguar cuánta energía se liberaría cuando el carbono y el oxígeno forman dióxido de carbono. El único problema es que las diferencias de masa son tan pequeñas que es muy difícil de realizarlo técnicamente.

Volvamos ahora al problema de si debemos agregar m_0c^2 a la energía cinética y decir de ahora en adelante que la energía total de un objeto es mc^2. Primero, si todavía podemos *ver* las partes componentes de masa en reposo m_0 dentro de M, entonces podríamos decir que alguna parte de la masa M del objeto compuesto es la masa de reposo mecánica de las partes componentes, otra parte es energía cinética de las partes y otra parte es energía potencial de las partes. Pero hemos descubierto en la naturaleza partículas de varios tipos que sufren reacciones como la que hemos tratado más arriba, en las cuales, con todo el estudio en el mundo, *no podemos ver las partes adentro*. Por ejemplo, cuando un mesón K se desintegra en dos piones, lo hace de acuerdo con la ley (16.11), pero la idea de que un K está hecho de $2\,\pi$ es una idea sin valor, ¡porque también se desintegra en $3\,\pi$!

Por lo tanto, tenemos una *nueva idea*: no necesitamos saber de qué están constituidas las cosas adentro, no podemos ni necesitamos identificar, dentro de una partícula, qué parte de la energía es energía de reposo de las partes componentes en las cuales se va a desintegrar. No es conveniente y a menudo no es posible separar la energía total mc^2 de un objeto en energía de reposo de las partes de adentro, energía cinética de las partes y energía potencial de las partes; en su lugar, sencillamente hablaremos de la *energía total* de la partícula. "Cambiamos el origen" de la energía agregando una constante m_0c^2 a todo, además decimos que la energía total de una partícula es la masa en movimiento por c^2, y cuando el objeto esté quieto, la energía es la masa en reposo por c^2.

Finalmente, encontramos que la velocidad v, el momentum P y la energía total E están relacionadas de una manera simple. Que la masa en movimiento a velocidad v es la masa m_0 en reposo dividida entre $\sqrt{1 - v^2/c^2}$, sorprendentemente, se usa muy raramente. En cambio, las siguientes reacciones se prueban fácilmente y resultan muy útiles,

$$E^2 - P^2c^2 = m_0^2c^4 \tag{16.13}$$

y

$$Pc = Ev/c. \tag{16.14}$$

17

Espacio-tiempo

17-1 La geometría del espacio-tiempo

La teoría de la relatividad nos muestra que las relaciones entre posiciones y tiempos medidos en un sistema coordenado o en otro no son lo que habíamos esperado en base a nuestras ideas intuitivas. Es muy importante que entendamos concienzudamente las relaciones entre espacio y tiempo implicadas por la transformación de Lorentz y, por lo tanto, vamos a considerar este asunto con más profundidad en este capítulo.

La transformación de Lorentz entre las posiciones y el tiempo (x, y, z, t) medidos por un observador "en reposo", y las coordenadas y el tiempo correspondientes (x', y', z', t') medidos dentro de una nave espacial "en movimiento", que se mueve con velocidad u, son

$$\begin{aligned} x' &= \frac{x - ut}{\sqrt{1 - u^2/c^2}}, \\ y' &= y, \\ z' &= z, \\ t' &= \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}}. \end{aligned} \tag{17.1}$$

Comparemos estas ecuaciones con la ecuación (11.5), que también relaciona medidas en dos sistemas, uno de los cuales, en este caso, está "rotado" respecto al otro:

$$\begin{aligned} x' &= x\cos\theta + y\operatorname{sen}\theta, \\ y' &= y\cos\theta - x\operatorname{sen}\theta, \\ z' &= z. \end{aligned} \tag{17.2}$$

En este caso particular, Pedro y Juan están midiendo con ejes que tienen un ángulo θ entre los ejes x' y x. En cada caso notamos que las cantidades "con prima" son "mezclas" de las "sin primas": la nueva x' es una mezcla de x y y, y la nueva y' también es una mezcla de x y y.

Una analogía es útil: cuando miramos un objeto hay una cosa evidente que podemos llamar el "ancho aparente" y otra que podemos llamar "profundidad". Pero los dos conceptos ancho y profundidad no son propiedades *fundamentales* del objeto, porque si nos apartamos y miramos el mismo objeto desde un ángulo diferente, obtenemos un ancho diferente y una profundidad diferente, y podemos desarrollar algunas fórmulas para calcular las nuevas a partir de las antiguas y de los ángulos comprendidos. Las ecuaciones (17.2) son estas fórmulas. Uno podría decir que una profundidad dada es una especie de mezcla de toda la profundidad y de todo el ancho. Si no pudiéramos movernos nunca y siempre viéramos un objeto dado desde una misma posición, todo este razonamiento no tendría objeto, veríamos siempre el "verdadero" ancho y la "verdadera" profundidad, los cuales aparecerían con cualidades bastante diferentes, porque uno aparece como un ángulo óptico subtendido y la otra comprende un enfocamiento de los ojos o aun intuición; parecerían ser cosas muy diferentes y nunca se mezclarían. Es porque *podemos* caminar que nos damos cuenta que la profundidad y el ancho son, de una manera u otra, sólo dos aspectos de una misma cosa.

¿No podemos mirar la transformación de Lorentz de la misma manera? Aquí también tenemos una mezcla —de posiciones y tiempo—. Una diferencia entre una medida de espacio y una medida de tiempo produce una nueva medida de espacio. En otras palabras, en las medidas de espacio hechas por un hombre,

hay mezclado un poco de tiempo, visto por otro. Nuestra analogía nos permite generar esta idea: la "realidad" de un objeto que estamos mirando es de alguna manera más grande (hablando en forma aproximada e intuitiva) que su "ancho" y su "profundidad", porque *dependen* de *qué manera* lo miramos; cuando nos movemos a una nueva posición, nuestra mente inmediatamente vuelve a calcular el ancho y la profundidad. Pero nuestra mente no recalcula inmediatamente las coordenadas y el tiempo cuando nos movemos a alta velocidad, porque no tenemos experiencia efectiva de ir casi tan rápido como la luz, para apreciar el hecho que el tiempo y el espacio son también de la misma naturaleza. Es como si estuviéramos condenados a mirar solamente el ancho de algo, sin poder mover nuestras cabezas suficientemente hacia un lado o hacia el otro; comprendemos ahora que, si pudiéramos, veríamos algo del tiempo del otro hombre —veríamos "detrás" un poquito, por decir así—.

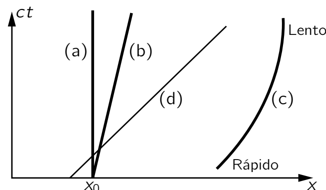


Figura 17-1. Tres trayectorias de una partícula en el espacio-tiempo: (a) una partícula en reposo en $x = x_0$; (b) una partícula que parte de $x = x_0$ y se mueve con velocidad constante; (c) una partícula que parte con alta velocidad, pero después se retarda; (d) un haz de luz.

Por lo tanto, vamos a tratar de pensar en objetos en un mundo de naturaleza nueva, de espacio y tiempo mezclados, en el mismo sentido que los objetos en nuestro mundo de espacio ordinario son reales y pueden ser mirados desde diferentes direcciones. Consideramos después que objetos que ocupan espacio y duran un cierto tiempo ocupan una especie de "burbuja" en un nuevo tipo de mundo y que miramos a esta "burbuja" desde diferentes puntos de vista cuando nos estamos moviendo a diferentes velocidades. Este nuevo mundo, esta entidad geométrica en la cual la "burbuja" existe al ocupar una posición y tomar una cierta cantidad de tiempo, se llama *espacio-tiempo*. Un punto dado $(x, y, z, t,)$ en el espacio-tiempo se llama *evento*. Imaginen, por ejemplo, que graficamos las posiciones x horizontalmente, y y z en otras dos direcciones, ambas "perpendiculares" entre sí y "perpendiculares" al papel (!) y el tiempo verticalmente. Ahora, ¿cómo se ve una partícula en movimiento en este gráfico? Si la partícula está en reposo, tiene una cierta x y a medida que el tiempo pasa tiene la misma x, la misma x, la misma x, de manera que su "trayectoria" es una línea paralela al eje t (figura 17-1a). Por el contrario, si se desplaza, entonces a medida que el tiempo transcurre x aumenta (figura 17-1b). De manera que, por ejemplo, una partícula, que empieza a desplazarse después va más lenta, debería tener un movimiento como el que se muestra en la figura 17-1c. Una partícula, en otras palabras, que es permanente y no se desintegra está representada por una línea en el espacio-tiempo. Una partícula que se desintegra estaría representada por una línea en forma de horqueta, porque se transformaría en otras dos cosas que partirían de ese punto.

¿Y qué pasa con la luz? La luz viaja con velocidad c y estaría representada por una línea con cierta pendiente fija (figura 17-1d).

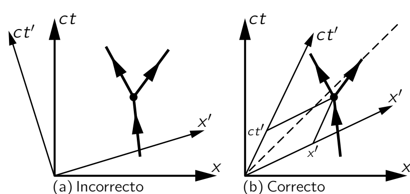


Figura 17-2. Dos vistas de una partícula desintegrándose.

Ahora bien, de acuerdo con nuestra nueva idea, si una partícula verifica un evento dado, por ejemplo, si se desintegrara súbitamente un cierto punto del espacio-tiempo en dos nuevas partículas que siguen nuevas trayectorias, y si este interesante evento ocurriera en un cierto valor de x y un cierto valor de t, entonces podríamos esperar que, siempre que esto tenga algún sentido, sencillamente tenemos que tomar un nuevo par de ejes y rotarlos y eso nos dará el nuevo t y la nueva x en nuestro sistema nuevo, como se muestra en la figura 17-2a. Pero esto está mal, porque la ecuación (17.1) no es exactamente la misma transformación matemática que la ecuación (17.2). Hay que hacer notar, por ejemplo, la diferencia de signo entre las dos y el hecho de que una esté escrita en términos de $\cos\theta$ y $\operatorname{sen}\theta$, mientras la otra está escrita con cantidades algebraicas. (Por supuesto, no es imposible que las cantidades algebraicas se pudieran escribir como coseno y seno, pero en realidad no se puede.) Pero de todas maneras, las dos expresiones *son* muy similares. Como vamos a ver, realmente no es posible pensar que el espacio-tiempo tenga una geometría real y ordinaria debido a esa diferencia de signo. En realidad, aunque no vamos a hacer énfasis en este punto, resulta que un hombre que se mueve tiene que usar un conjunto de ejes que están igualmente inclinados respecto al rayo de luz, usando un tipo de proyección especial paralela a los ejes x' y t', para su x' y t' como se muestra en la figura 17-2b. No vamos a trabajar con la geometría porque no ayuda mucho; es más fácil trabajar con las ecuaciones.

17-2 Intervalos de espacio-tiempo

Aunque la geometría del espacio-tiempo no es euclideana en el sentido ordinario, *hay* una geometría que es muy similar, pero peculiar en ciertos aspectos. Si esta idea de geometría está correcta deberían existir algunas funciones de las coordenadas y el tiempo que sean independientes del sistema de coordenadas. Si, por ejemplo, en rotaciones ordinarias, tomamos dos puntos, uno al origen para mayor sencillez, y el otro en cualquier otra parte, ambos sistemas tendrían el mismo origen y la distancia desde aquí hasta el otro punto es la misma para ambos. Ésta es una propiedad que es independiente de la manera particular de medirla. El cuadrado de la distancia es $x^2 + y^2 + z^2$. ¿Y qué hay ahora en el espacio-tiempo? No es difícil demostrar que aquí tenemos también algo que se mantiene igual, a saber, la combinación $c^2t^2 - x^2 - y^2 - z^2$ es la misma antes y después de la transformación:

$$c^2t'^2 - x'^2 - y'^2 - z'^2 = c^2t^2 - x^2 - y^2 - z^2. \tag{17.3}$$

Esta cantidad es, por lo tanto, algo que, lo mismo que la distancia, es "real" en cierto sentido; se llama *intervalo* entre dos puntos del espacio-tiempo, uno de los cuales está, en este caso, en el origen. (Realmente, por supuesto, es el intervalo al cuadrado, así como $x^2 + y^2 + z^2$ es la distancia al cuadrado.) Le damos un nombre diferente porque está en una geometría diferente, pero lo interesante es que sólo algunos signos están cambiados y que hay una c.

Deshagámonos de la c; es un absurdo si vamos a tener un espacio maravilloso con las x y y que pueden ser intercambiadas. Una de las confusiones que podría ser causada por alguien sin experiencia sería medir anchos, digamos, mediante el ángulo subtendido por el ojo y medir profundidades de una manera diferente, como el esfuerzo muscular necesario para enfocarlos, de manera que las profundidades estarían medidas en pies y los anchos en metros. Entonces uno obtendría un enredo enormemente complicado de ecuaciones al hacer transformaciones como la (17.2), y no sería capaz de ver la claridad y sencillez de la cosa por una simple razón técnica, que la misma cosa se está midiendo en dos unidades diferentes. Ahora bien, en las ecuaciones (17.1) y (17.3) la naturaleza nos está diciendo que el tiempo y el espacio son equivalentes; el tiempo se transforma en espacio; *deben ser medidos en las mismas unidades*. ¿Qué distancia es un "segundo"? Es fácil de calcular a partir de (17.3). Es 3×10^8 metros, *la distancia que la luz recorrería en un segundo*. En otras palabras, si midiéramos todas las distancias y tiempos en las mismas unidades, segundos, entonces nuestra unidad de distancia sería 3×10^8 metros y las ecuaciones serían más sencillas. Otra manera de hacer las unidades iguales es que midiéramos el tiempo en metros. ¿Qué es un metro de tiempo? Un metro de tiempo es el tiempo que demora la luz en avanzar un metro y, por lo tanto, es $1/3 \times 10^{-8}$ segundos, o ¡3 300 millonésimas de segundo! Desearíamos, en otras palabras, poner nuestras ecuaciones en un sistema de unidades en el cual $c = 1$. Si el tiempo y el espacio están medidos en las mismas unidades, como se sugirió, evidentemente las ecuaciones quedan mucho más simplificadas. Ellas son

$$\begin{aligned} x' &= \frac{x - ut}{\sqrt{1 - u^2}}, \\ y' &= y, \\ z' &= z, \\ t' &= \frac{t - ux}{\sqrt{1 - u^2}}. \end{aligned} \tag{17.4}$$

$$t'^2 - x'^2 - y'^2 - z'^2 = t^2 - x^2 - y^2 - z^2. \tag{17.5}$$

Si alguna vez estamos poco seguros o "asustados" de que una vez que tengamos este sistema con $c = 1$ no vamos a poder obtener nuestras ecuaciones correctamente de nuevo, la respuesta es justamente lo opuesto. Es mucho más fácil recordarlas sin la c, y es muy fácil poner de nuevo la c examinando las dimensiones. Por ejemplo, en $\sqrt{1 - u^2}$, sabemos que no podemos restar el cuadrado de una

velocidad, que tiene unidades, del número puro 1, de modo que sabemos que debemos dividir u^2 por c^2 de manera de que quede sin unidades, y así sigue todo.

La diferencia entre el espacio-tiempo y el espacio ordinario, y el carácter de un intervalo relacionado con la distancia, es muy interesante. De acuerdo con la fórmula (17.5), si consideramos un punto que en un sistema de coordenadas dado tenía un tiempo cero, y sólo espacio, el cuadrado del intervalo sería negativo y tendríamos un intervalo imaginario, la raíz cuadrada de un número negativo. Los intervalos pueden ser reales o imaginarios en la teoría. Los cuadrados de los intervalos pueden ser positivos o negativos, a diferencia de la distancia que tiene un cuadrado positivo. Cuando un intervalo es imaginario, decimos que los dos puntos tienen un *intervalo de tipo espacio* entre ellos (en vez de imaginario), porque el intervalo es más como espacio que como tiempo. Por otro lado, si dos objetos están en el mismo lugar en un sistema de coordenadas dado, pero difieren sólo en el tiempo, entonces el cuadrado del tiempo es positivo y las distancias son cero y el intervalo al cuadrado es positivo; éste se llama *intervalo tipo tiempo*. En nuestro diagrama de espacio-tiempo, por lo tanto, tendremos más o menos la siguiente representación: a 45° hay dos líneas (realmente, en cuatro dimensiones éstas serían "conos", llamados conos de luz) y los puntos de estas líneas están todos a un intervalo cero desde el origen. A cualquier parte que la luz vaya desde un punto dado está siempre separada de él por un intervalo cero, como se deduce de la ecuación (17.5). A propósito, acabamos de probar que si la luz viaja con velocidad c en un sistema, viaja con velocidad c en otro, porque si el intervalo es el mismo en ambos sistemas, en otras palabras, cero en uno y cero en el otro, entonces decir que la velocidad de propagación de la luz es invariante es lo mismo que decir que el intervalo es cero.

17-3 Pasado, presente y futuro

La región espacio-tiempo que rodea un punto de espacio-tiempo dado puede ser separada en tres regiones, como se muestra en la figura 17-3. En una región tenemos intervalos de tipo espacio, y en dos regiones intervalos de tipo tiempo. Físicamente, estas tres regiones en las cuales se divide el espacio-tiempo alrededor de un punto dado tienen una interesante relación física con ese punto: un objeto físico o una señal puede llegar desde un punto en la región 2 a un evento O moviéndose con una velocidad menor que la velocidad de la luz.

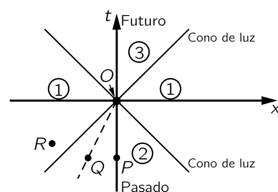


Figura 17-3. La región espacio-tiempo que rodea un punto en el origen.

Por lo tanto, eventos en esta región pueden afectar el punto O, pueden tener una influencia sobre él desde el pasado. En verdad, por supuesto, un objeto en P sobre el eje t negativo está precisamente en el "pasado" con respecto a O; es el mismo punto espacio que O, sólo que más temprano. Lo que sucedió ahí entonces afecta a O ahora. (Desgraciadamente, la vida es así.) Otro objeto en Q puede llegar a O moviéndose con una cierta velocidad menor que c, de manera que si este objeto estuviera en una nave espacial y moviéndose, sería también el pasado del mismo punto de espacio. O sea, en otro sistema de coordenadas, el eje del tiempo podría pasar por O y Q. Luego, todos los puntos de la región 2 están en el pasado de O, y cualquier cosa que suceda en esta región puede *afectar* a O. Por lo tanto, la región 2 se llama a veces *pasado afectable*, o pasado que puede afectar; es el lugar geométrico de todos los eventos que pueden afectar al punto O de alguna manera.

La región 3, por otro lado, es una región que nosotros podemos afectar *desde* O, podemos "golpear" cosas disparando "balas" a velocidades menores que c. Luego es el mundo cuyo futuro puede ser afectado por nosotros y lo podemos llamar el *futuro afectable*. Ahora bien, lo interesante acerca de todo el resto del espacio-tiempo, es decir, la región 1, es que no podemos afectarla ahora *desde* O, ni puede ella afectarnos a nosotros ahora *en* O, porque nada puede ir más rápido que la luz. Por supuesto, lo que sucede en R *puede* afectarnos *más tarde*; es decir, si el Sol está explotando "en este mismo momento", nos toma ocho minutos antes que sepamos de ello, y no nos puede afectar de ninguna manera antes de entonces.

Lo que entendemos por “en este mismo momento” es algo misterioso que no podemos definir y no podemos afectar, pero nos puede afectar más tarde, o podríamos haber afectado si hubiéramos hecho algo con suficiente anterioridad en el pasado. Cuando miramos a la estrella Alfa Centauri, la vemos como era hace cuatro años; podríamos preguntarnos cómo es “ahora”. “Ahora” significa al mismo tiempo desde nuestro sistema de coordenadas especial. Podemos ver Alfa Centauri solamente mediante la luz que ha venido de nuestro pasado, hasta hace cuatro años, pero no sabemos lo que está haciendo “ahora”; van a pasar cuatro años antes que lo que está haciendo “ahora” pueda afectarnos. Alfa Centauri “ahora” es una idea o concepto de nuestra mente; no es algo que sea realmente definible físicamente en este momento, porque tenemos que esperar para observarlo, no podemos siquiera definirlo “ahora” mismo. Además, el “ahora” depende del sistema de coordenadas. Si, por ejemplo, Alfa Centauri se estuviera moviendo, un observador allí no estaría de acuerdo con nosotros porque pondría sus ejes formando un ángulo, y su “ahora” sería un tiempo “diferente”. Ya hemos hablado del hecho de que la simultaneidad no es una cosa única.

Hay adivinos, o personas que nos dicen que pueden conocer el futuro y hay muchas hermosas historias acerca del hombre que súbitamente descubre que tiene conocimiento del futuro afectable. Bueno, hay muchas paradojas producidas por eso, porque si sabemos que algo va a suceder, podemos asegurarnos que lo evitaremos haciendo lo necesario en el momento preciso, etc. Pero en realidad, no hay ningún adivino que pueda decirnos ni siquiera el *¡presente!*. No hay nadie que pueda decirnos qué está sucediendo realmente en este mismo momento, a una distancia razonable, porque no es observable. Podríamos hacernos esta pregunta que dejamos al estudiante que trate de contestar: ¿Se produciría alguna paradoja si súbitamente se hiciera posible conocer cosas que están en intervalos de tipo espacio en la región 1?

17-4 Más acerca de los vectores de cuatro dimensiones

Volvamos ahora a nuestras consideraciones sobre la analogía de la transformación de Lorentz y las rotaciones de ejes espaciales. Hemos aprendido la utilidad de juntar otras cantidades que tienen las mismas propiedades de transformación que las coordenadas, para formar lo que llamamos *vectores*, líneas dirigidas. En el caso de rotaciones ordinarias, hay muchas cantidades que se transforman de la misma manera que x, y y z con una rotación: por ejemplo, la velocidad tiene tres componentes, una x, una y y una z; cuando se las observa en un sistema diferente de coordenadas, ninguna de las componentes es la misma, sino que se han transformado a un nuevo valor. Pero, de alguna manera, la velocidad “misma” tiene mayor realidad que cualquiera de sus componentes particulares, y la representamos por un segmento dirigido.

Preguntamos, por lo tanto: ¿Es verdad o no que existen cantidades que se transforman o que están relacionadas en un sistema en movimiento y en un sistema inmóvil de la misma manera que x, y, z y t? Por nuestra experiencia con vectores sabemos que tres de las cantidades como x, y, z constituirían las tres componentes de un vector espacial ordinario, pero la cuarta cantidad parecería un simple escalar en una rotación espacial, porque no cambia mientras no vayamos a un sistema de coordenadas en movimiento. ¿Es posible, entonces, asociar con algunos de nuestros “vectores de dimensión tres” un cuarto objeto, que llamaríamos la “componente tiempo”, de tal manera que los cuatro objetos juntos “roten” de la misma manera que la posición y el tiempo en el espacio-tiempo? Vamos a mostrar ahora que hay ciertamente, por lo menos, un objeto así (hay muchos en realidad): *las tres componentes del momentum y la energía como la componente tiempo, se transforman conjuntamente* para hacer lo que llamamos un “vector de dimensión cuatro”. Para demostrar esto, ya que es muy inconveniente tener que escribir c en todas partes, usaremos el mismo truco referente a unidades de energía, de masa y de momentum que usamos en la ecuación (17.4). Energía y masa, por ejemplo, difieren sólo en un factor c^2 lo que es simplemente una cosa de unidades, de manera que podemos decir que la energía *es* la masa. En vez de

tener que escribir la c^2 ponemos $E = m$, y entonces, por supuesto, si hay algún contratiempo, pondremos de nuevo c en cantidad apropiada de manera que las unidades se corrijan en la última ecuación, pero no en las intermedias.

Luego, nuestras ecuaciones para la energía y la cantidad de movimiento son

$$\begin{aligned} E &= m = m_0/\sqrt{1-v^2}, \\ \boldsymbol{p} &= m\boldsymbol{v} = m_0\boldsymbol{v}/\sqrt{1-v^2}. \end{aligned} \tag{17.6}$$

También en estas unidades tenemos

$$E^2 - p^2 = m_0^2. \tag{17.7}$$

Por ejemplo, si medimos la energía en electrón volts, ¿qué significaría una masa de un electrón volt? Significa la masa cuya energía de reposo es un electrón volt, es decir, m_0c^2 es un electron volt. Por ejemplo, la masa en reposo de un electrón es 0.511×10^6 eV.

Ahora bien, ¿cómo se verían el momentum y la energía en un nuevo sistema de coordenadas? Para averiguarlo, deberemos transformar la ecuación (17.6), lo que podemos hacer porque sabemos cómo se transforman las velocidades. Supongamos que mientras lo medimos, un objeto tenga una velocidad v, pero miramos el mismo objeto desde el punto de vista de una nave espacial que a su vez se está moviendo con una velocidad u, y en ese sistema usamos una prima para designar la cosa correspondiente. Para simplificar las cosas al principio, vamos a considerar el caso en que la velocidad v está en la dirección de u. (Después podemos hacer el caso más general.) ¿Cuál es v', la velocidad vista desde la nave espacial? Es la velocidad compuesta, la "diferencia" entre v y u. Mediante la ley que dedujimos anteriormente

$$v' = \frac{v-u}{1-uv}. \tag{17.8}$$

Ahora calculamos la nueva energía E', la energía tal como la vería el individuo de la nave espacial. Él usaría la misma masa en reposo, por cierto, pero usaría v' para la velocidad. Lo que tenemos que hacer es elevar al cuadrado v', restarlo de uno, extraer la raíz cuadrada y tomar la inversa:

$$\begin{aligned} v'^2 &= \frac{v^2 - 2uv + u^2}{1 - 2uv + u^2v^2}, \\ 1 - v'^2 &= \frac{1 - 2uv + u^2v^2 - v^2 + 2uv - u^2}{1 - 2uv + u^2v^2}, \\ &= \frac{1 - v^2 - u^2 + u^2v^2}{1 - 2uv + u^2v^2}, \\ &= \frac{(1-v^2)(1-u^2)}{(1-uv)^2}. \end{aligned}$$

Por lo tanto,

$$\frac{1}{\sqrt{1-v'^2}} = \frac{1-uv}{\sqrt{1-v^2}\sqrt{1-u^2}}. \tag{17.9}$$

La energía E' es entonces simplemente m_0 por la expresión anterior. Pero deseamos expresar la energía en términos de la energía y momentum sin primas, y notamos que

$$E' = \frac{m_0 - m_0uv}{\sqrt{1-v^2}\sqrt{1-u^2}} = \frac{(m_0/\sqrt{1-v^2}) - (m_0v/\sqrt{1-v^2})u}{\sqrt{1-u^2}},$$

o

$$E' = \frac{E - up_x}{\sqrt{1-u^2}}, \tag{17.10}$$

que reconocemos que tiene exactamente la misma forma que

$$t' = \frac{t - ux}{\sqrt{1-u^2}}.$$

A continuación, debemos encontrar la nueva cantidad de movimiento p'_x. Esto es simplemente la energía E multiplicada por v', y es expresada también en forma simple en términos de E y p:

$$p'_x = E'v' = \frac{m_0(1-uv)}{\sqrt{1-v^2}\sqrt{1-u^2}} \cdot \frac{v-u}{(1-uv)} = \frac{m_0 v - m_0 u}{\sqrt{1-v^2}\sqrt{1-u^2}}.$$

Así

$$p'_x = \frac{p_x - uE}{\sqrt{1-u^2}}, \tag{17.11}$$

que reconocemos que tiene precisamente la misma forma que

$$x' = \frac{x-ut}{\sqrt{1-u^2}}.$$

Luego, las transformaciones para la nueva energía y momentum en términos de la energía y el momentum antiguos son exactamente lo mismo que las transformaciones para t' en términos de t y x, y de x' en términos de x y t: todo lo que tenemos que hacer es, cada vez que vemos t en (17.4) sustituirla por E, y cada vez que veamos x sustituirla por p_x, y entonces las ecuaciones (17.4) van a ser iguales a las ecuaciones (17.10) y (17.11). Esto implicaría, si todo resulta bien, una regla adicional: $p'_y = p_y$ y $p'_z = p_z$. Para probar esto, se necesitaría volver atrás y estudiar el caso del movimiento hacia arriba y hacia abajo. Realmente, nosotros estudiamos el caso del movimiento hacia arriba y hacia abajo en el último capítulo. Analizamos un choque complicado y notamos, en realidad, que el momentum transversal *no* se cambia cuando se observa desde un sistema en movimiento: así que ya hemos verificado que $p'_y = p_y$ y $p'_z = p_z$. La transformación completa es entonces

$$\begin{aligned} p'_x &= \frac{p_x - uE}{\sqrt{1-u^2}}, \\ p'_y &= p_y, \\ p'_z &= p_z, \\ E' &= \frac{E - up_x}{\sqrt{1-u^2}}. \end{aligned} \tag{17.12}$$

En estas transformaciones, por lo tanto, hemos descubierto cuatro cantidades que se transforman como x, y y t, y que llamamos el *vector de dimensión cuatro momentum.* Como el momentum es un vector de dimensión cuatro, se le puede representar en un diagrama espacio-tiempo de una partícula en movimiento como una "flecha" tangente a la trayectoria, como se muestra en la figura 17-4. Esta flecha tiene una componente temporal igual a la energía y sus componentes espaciales representan su vector de dimensión tres momentum; esta flecha es más "real" que la energía o el momentum, porque éstos dependen precisamente de la manera en que miramos el diagrama.

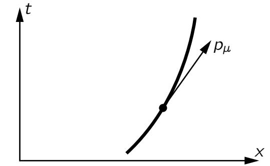


Figura 17-4. El vector de dimensión cuatro momentum de una partícula.

17-5 Álgebra de vectores de cuatro dimensiones

La notación para los vectores de cuatro dimensiones es diferente que para los vectores de dimensión tres. En el caso de vectores de dimensión tres, si fuéramos a hablar del vector de dimensión tres momentum ordinario lo escribiríamos $\boldsymbol{p}$. Si quisiéramos ser más específicos, podríamos decir que tiene tres componentes que son, para los ejes en consideración, p_x, p_y y p_z, o simplemente referirnos a una componente general p_i, y decir que i puede ser x, y o z y que éstas son las tres componentes; esto es, imagínense que i es cualquiera de estas tres direcciones x, y o z. La notación que usamos para vectores de dimensión cuatro es análoga a ésta: escribimos p_μ para vectores de dimensión cuatro y μ representa las *cuatro* direcciones posibles t, x, y o z.

Podríamos, por supuesto, usar cualquier notación que quisiéramos, no se rían de las notaciones; invéntenlas, son muy poderosas. De hecho, la matemática es,

en gran parte, invención de notaciones mejores. Todo el concepto de vector de dimensión cuatro es, de hecho, un mejoramiento en la notación de manera que las transformaciones se puedan recordar con facilidad. A_μ es un vector de dimensión cuatro general, pero para el caso especial del momentum, el p_t se identifica como la energía, p_x es el momentum en la dirección x, P_y en la dirección de y y p_z en la dirección z. Para sumar vectores de dimensión cuatro sumamos las componentes correspondientes.

Si hay una ecuación entre vectores de dimensión cuatro, entonces la ecuación es válida para *cada componente*. Por ejemplo, si la ley de conservación del vector de dimensión tres momentum ha de ser válida en choques de partículas, es decir, si la suma de los momentos para un gran número de partículas interactuando o chocando debe ser constante, eso tiene que significar que todos los momentos en la dirección x, en la dirección y y en la dirección z, para todas las partículas, deben ser constantes. Esta ley sola no sería posible en relatividad porque está *incompleta*; es como hablar sólo de dos de las componentes de un vector de dimensión tres. Es incompleta porque si rotamos los ejes mezclamos las diferentes componentes, así que debemos incluir las tres componentes en nuestra ley. Luego en relatividad debemos completar la ley de conservación del momentum, extendiéndola de modo de incluir la componente *tiempo*. Es *absolutamente necesario* que ésta vaya con las otras tres, o no puede haber invariancia relativista. La *conservación de energía* es la cuarta ecuación que va con la conservación del momentum para lograr una relación entre vectores de dimensión cuatro válida en la geometría del espacio y el tiempo. Así pues, la ley de conservación de energía y momentum en notación con vectores de dimensión cuatro es

$$\sum_{\substack{\text{partículas}\\ \text{entrantes}}} p_\mu = \sum_{\substack{\text{partículas}\\ \text{salientes}}} p_\mu \tag{17.13}$$

o, en una notación ligeramente diferente

$$\sum_i p_{i\mu} = \sum_j p_{j\mu}, \tag{17.14}$$

donde $i = 1, 2, \ldots$ se refiere a las partículas que van a chocar, $j = 1, 2, \ldots$ se refiere a las partículas que acaban de chocar y $\mu = x$, y, z o t. Ustedes dirán "¿en qué ejes?" Da lo mismo. La ley es válida para cada componente usando *cualquier* eje.

En análisis vectorial discutimos otra cosa, el producto escalar de dos vectores. Consideremos lo correspondiente en el espacio-tiempo. En una rotación ordinaria descubrimos que había una cantidad que no cambiaba $x^2 + y^2 + z^2$. En cuatro dimensiones la cantidad correspondiente es $t^2 - x^2 - y^2 - z^2$ [ecuación (17.3)]. ¿Cómo podemos escribir eso? Una manera sería escribir algún tipo de cosa en dimensión cuatro con un punto cuadrado en el medio, tal como $A_\mu \boxdot B_\mu$; una de las notaciones que se usan realmente es

$$\sum_\mu{}' A_\mu A_\mu = A_t^2 - A_x^2 - A_y^2 - A_z^2. \tag{17.15}$$

La prima en $\sum$ significa que el primer término, el término "temporal", es positivo pero los otros tres términos tienen signo negativo. Esta cantidad, entonces, va a ser la misma en cualquier sistema de coordenadas, y podríamos llamarla cuadrado de la longitud del vector de dimensión cuatro. Por ejemplo, ¿cuál es el cuadrado de la longitud del vector de dimensión cuatro momentum de una partícula? Sería igual a $p_t^2 - p_x^2 - p_y^2 - p_z^2$, o en otras palabras $E^2 - p^2$, porque sabemos que p_t es E. ¿Qué es $E^2 - p^2$? Debe ser algo que es lo mismo en todo sistema de coordenadas. En particular, debe ser lo mismo para un sistema de coordenadas que se mueve junto con la partícula, en el cual la partícula está en reposo. Si la partícula está en reposo no tiene momentum. Luego en este sistema de coordenadas es su energía solamente, que es lo mismo que su masa en reposo. Por lo tanto, $E^2 - p^2 = m_0^2$. Vemos así que el cuadrado de la longitud de este vector, el vector de dimensión cuatro momentum, es igual a m_0^2 .

Del cuadrado de un vector, podemos proseguir e inventar el "producto escalar", el producto que es un escalar: si a_μ es un vector de dimensión cuatro y b_μ es otro vector de dimensión cuatro, el producto escalar es

$$\sum\nolimits' a_\mu b_\mu = a_t b_t - a_x b_x - a_y b_y - a_z b_z. \tag{17.16}$$

Es el mismo en todos los sistemas de coordenadas.

Finalmente, mencionaremos algunas cosas cuya masa en reposo m_0 es cero. Un fotón de luz, por ejemplo. Un fotón es como una partícula, en el sentido que lleva energía y momentum. La energía de un fotón es una cierta constante, llamada constante de Planck, multiplicada por la frecuencia del fotón: $E = h\nu$. Este fotón también lleva momentum, y el momentum de un fotón (o de cualquier partícula, en realidad) es h dividida entre la longitud de onda: $p = h/\lambda$. Pero para un fotón, hay una relación bien definida entre la frecuencia y la longitud de onda: $\nu = c/\lambda$. (El número de ondas por segundo multiplicado por la longitud de onda de cada una, es la distancia que la luz recorre en un segundo que es, por supuesto, c). Vemos así inmediatamente que la energía de un fotón debe ser el momentum multiplicado por c, o si $c = 1$, *la energía y el momentum son iguales*. Es decir, la masa en reposo es cero. Observemos esto de nuevo; es bastante curioso. Si tenemos una partícula de masa en reposo cero, ¿qué pasa cuando se detiene? *¡Nunca se detiene!* Siempre va a velocidad c. La fórmula corriente para la energía es $m_0/\sqrt{1 - v^2}$. ¿Podemos decir que $m_0 = 0$ y $v = 1$, de manera que la energía sea cero? *No podemos* decir que es cero; el fotón realmente puede (y debe) tener energía aunque no tenga masa en reposo; pero ¡la posee yendo perpetuamente a la velocidad de la luz!

También sabemos que el momentum de cualquier partícula es igual a su energía total por su velocidad: si $c = 1$, $p = vE$ o, en unidades ordinarias, $p = vE/c^2$. Para cualquier partícula que se mueve a la velocidad de la luz, $p = E$ si $c = 1$. Las fórmulas para la energía de un fotón vistas desde un sistema móvil están, por supuesto, dadas por la ecuación (17.12), pero debemos sustituir el momentum por la energía multiplicada por c (o por 1 en este caso). Las diferentes energías después de la transformación significan que hay frecuencias diferentes. Esto se llama efecto Doppler, y se puede calcular fácilmente a partir de la ecuación (17.12), usando también $E = p$ y $E = hv$.

Como dijo Minkowski, "el espacio en sí y el tiempo en sí se van a sumergir en meras sombras, y sólo una cierta unión entre ellos va a sobrevivir".

18

Rotación en dos dimensiones

18-1 El centro de masa

En los capítulos anteriores hemos estado estudiando la mecánica de los puntos o pequeñas partículas cuya estructura interna no nos preocupa. En los próximos capítulos vamos a estudiar la aplicación de las leyes de Newton a cosas más complicadas. Cuando el mundo se pone más complicado, se pone también más interesante y vamos a encontrar que los fenómenos asociados con la mecánica de un objeto más complejo que un punto son bastante sorprendentes. Por supuesto, estos fenómenos encierran solamente combinaciones de las leyes de Newton, pero a veces es difícil de creer que solamente $F = ma$ está en juego.

Los objetos más complicados con que tratamos pueden ser de diferentes tipos: agua corriendo, galaxias arremolinándose, etc. El objeto "complicado" más simple de analizar, al principio, es lo que llamamos un *cuerpos rígidos*, un objeto sólido que está rotando a medida que se mueve. Sin embargo, aun un objeto tan simple puede tener el movimiento más complejo, y por lo tanto vamos a considerar primero los aspectos más simples de este movimiento, en el cual un cuerpo extenso rota alrededor de un *eje fijo*. Un punto dado en ese cuerpo se mueve entonces en un plano perpendicular a este eje. Esta rotación de un cuerpo alrededor de un eje fijo se llama *rotación plana* o rotación en dos dimensiones. Después generalizaremos el resultado a tres dimensiones, pero al hacerlo vamos a encontrar que, a diferencia del caso de la mecánica ordinaria de partículas, las rotaciones son sutiles y difíciles de entender a menos que primero obtengamos un sólido fundamento en dos dimensiones.

El primer teorema interesante acerca del movimiento de objetos complicados se puede observar en acción si tiramos al aire un objeto hecho de muchos bloques y palos unidos con cuerdas. Por supuesto sabemos que sigue una parábola porque lo estudiamos así para una partícula. Pero nuestro objeto ahora *no* es una partícula; se bambolea y se agita, etc. Sin embargo, sigue una parábola; uno lo puede ver. ¿*Qué* es lo que sigue una parábola? Ciertamente no el punto en la esquina del bloque, porque ése está bamboleándose; tampoco es el extremo del palo, o la mitad del palo, o la mitad del bloque. Pero *algo* sigue una parábola, hay un "centro" efectivo que se mueve en una parábola. Así pues, nuestro primer teorema sobre objetos complicados es demostrar que *existe* una posición media que sigue una parábola que es definible matemáticamente, pero que no es necesariamente un punto del material mismo. Se llama teorema del centro de masa y la demostración es como sigue.

Podemos considerar cualquier objeto como constituido por muchas partículas chicas, los átomos, con diversas fuerzas entre ellas. Sea i el índice que define a una de las partículas. (Hay millones de ellas, de manera que i llega a 10^{23} o algo por el estilo.) Entonces la fuerza sobre la partícula i es, por supuesto, la masa por la aceleración de esa partícula:

$$\boldsymbol{F}_i = m_i(d^2\boldsymbol{r}_i/dt^2). \tag{18.1}$$

En los próximos capítulos nuestros objetos en movimiento van a ser objetos en los cuales todas las partes se están moviendo a velocidades mucho menores que la velocidad de la luz, y vamos a usar las aproximaciones no relativistas para todas las cantidades. En estas circunstancias, la masa es constante, luego

$$\boldsymbol{F}_i = d^2(m_i\boldsymbol{r}_i)/dt^2. \tag{18.2}$$

Si ahora sumamos la fuerza sobre todas las partículas, es decir, si tomamos la suma de todas las $\boldsymbol{F}_i$ para todos los índices diferentes, obtenemos la fuerza total $\boldsymbol{F}$. En el otro miembro de la ecuación, obtenemos lo mismo que si hubiéramos sumado antes de derivar:

$$\sum_i \boldsymbol{F}_i = \boldsymbol{F} = \frac{d^2(\sum_i m_i \boldsymbol{r}_i)}{dt^2}. \tag{18.3}$$

Por lo tanto, la fuerza total es la segunda derivada de la suma de las masas multiplicadas por sus posiciones.

Ahora bien, la fuerza total en todas las partículas es la misma que la fuerza *externa*. ¿Por qué? A pesar de que hay toda clase de fuerzas sobre las partículas debido a las cuerdas, los bamboleos, los tirones, los empujones y las fuerzas atómicas y quién sabe que más y que tenemos que sumarlas, nos salva la tercera ley de Newton. Entre dos partículas cualesquiera, la acción y la reacción son iguales, de manera que cuando sumamos todas las ecuaciones, si dos partículas tienen fuerzas entre ellas, éstas se anulan en la suma; luego, el resultado neto es solamente las fuerzas que surgen de otras partículas que no están incluidas en el objeto cualquiera sobre el cual decidamos sumar. Por lo tanto, si la ecuación (18.3) es la suma sobre un cierto número de partículas, que en conjunto se llama "el objeto", la fuerza *externa* sobre el objeto total es igual a la suma de *todas* las fuerzas sobre todas sus partículas constituyentes.

Ahora sería bonito si pudiéramos escribir la ecuación (18.3) como la masa total multiplicada por alguna aceleración. Podemos. Digamos que M es la suma de todas las masas, es decir, la masa total. Entonces, si *definimos* un cierto vector $\boldsymbol{R}$

$$\boldsymbol{R} = \sum_i m_i \boldsymbol{r}_i / M, \tag{18.4}$$

como la ecuación (18.3), va a ser simplemente

$$\boldsymbol{F} = d^2(M\boldsymbol{R})/dt^2 = M(d^2\boldsymbol{R}/dt^2), \tag{18.5}$$

ya que M es constante. Luego, encontramos que la fuerza externa es la masa total multiplicada por la aceleración de un punto imaginario cuya ubicación es $\boldsymbol{R}$. Este punto se llama *centro de masa* del cuerpo. Es un punto que está por ahí en el "medio" del objeto, una especie de $\boldsymbol{r}$ promedio en el cual los diferentes $\boldsymbol{r}_i$ tienen peso o importancia proporcional a las masas.

Vamos a discutir este importante teorema con más detalle en un capítulo posterior y, por lo tanto, vamos a limitar nuestras observaciones a dos puntos: primero, si las fuerzas externas son cero, si el objeto estuviera flotando en el espacio vacío, podría arremolinarse, tambalearse y torcerse y hacer todo tipo de cosas. Pero el *centro de masa*, esta posición calculada, inventada artificialmente, en alguna parte en el medio, *se va a mover con velocidad constante*. En particular, si estaba inicialmente en reposo, se va a mantener en reposo. Así pues, si tenemos algo como una caja, a lo mejor una nave espacial con gente dentro, y calculamos la posición del centro de masa y encontramos que está en reposo, entonces el centro de masa va a continuar en reposo si ninguna fuerza externa está actuando en la caja. Por supuesto que la nave espacial se puede mover un poco en el espacio, pero eso es solamente porque la gente está caminando hacia atrás y hacia adelante adentro; cuando uno camina hacia el frente, la nave va hacia atrás de manera de mantener la posición promedio de todas las masas exactamente en el mismo lugar.

¿Es, por lo tanto, la propulsión de los cohetes absolutamente imposible porque no podemos mover el centro de masa? No; pero, por supuesto, encontramos que para propulsar una parte interesante del cohete, una parte sin importancia debe ser botada. En otras palabras, si partimos con un cohete a velocidad cero y escupimos un poco de gas desde la parte de atrás, entonces esta pequeña pompa de gas va hacia un lado mientras que la nave cohete va hacia el otro, pero el centro de masa está todavía exactamente donde estaba antes. De manera que

simplemente movemos la parte en la cual estamos interesados a expensas de la parte en la que no estamos interesados.

El segundo punto acerca del centro de masa, que es la causa por la cual lo introducimos en nuestra discusión en este momento, es que se puede tratar separadamente de los movimientos “internos” de un objeto y, por lo tanto, se le puede ignorar en nuestra discusión de la rotación.

18-2 Rotación de un cuerpo rígido

Discutamos ahora las rotaciones. Por supuesto, que un objeto común no rota simplemente, se bambolea, se sacude y se curva, de manera que para simplificar las cosas vamos a discutir el movimiento de un objeto ideal no existente que llamamos *cuerpo rígido*. Esto significa un objeto en el cual las fuerzas entre los átomos son tan fuertes y de tal carácter que las pequeñas fuerzas que se necesitan para moverlo no lo curvan. Su forma se mantiene esencialmente la misma cuando se mueve. Si queremos estudiar el movimiento de tal cuerpo y aceptamos ignorar el movimiento de su centro de masas, le queda una sola cosa por hacer, esto es, *girar*. Tenemos que describir esto. ¿Cómo? Supongamos que hay una línea cualquiera en el cuerpo que se mantiene fija (a lo mejor incluye al centro de masas o a lo mejor no), y el cuerpo está girando alrededor de esta línea particular como eje. ¿Cómo definimos la rotación? Eso es bastante fácil, porque si marcamos un punto en cualquier parte del objeto, en cualquier parte menos en el eje, podemos decir siempre exactamente dónde está el objeto, con sólo saber dónde ha ido este punto. Lo único que se necesita para describir la posición de ese punto es un *ángulo*. De manera que la rotación consiste en el análisis y estudio de la variación de cierto ángulo con el tiempo.

Para estudiar la rotación observamos el ángulo en el cual ha rotado el cuerpo. Por supuesto, que no nos estamos refiriendo a ningún ángulo particular *dentro* del objeto mismo; no es que dibujemos algún ángulo *en* el objeto. Estamos hablando del *cambio angular de la posición* de todo el objeto, de un instante al siguiente.

Primero, estudiemos la cinemática de las rotaciones. El ángulo va a cambiar con el tiempo y de la misma manera que hablábamos de posición y velocidad en una dimensión, podemos hablar acerca de posiciones angulares y velocidades angulares en la rotación plana. De hecho, hay una relación muy interesante entre la rotación en dos dimensiones y el desplazamiento en una dimensión, en la cual casi toda cantidad tiene su análogo. Primero tenemos el ángulo θ que define en *cuánto* ha girado el cuerpo; esto remplaza la distancia s, que define en cuánto se ha *desplazado*. De la misma manera, tenemos una velocidad de rotación, $\omega = d\theta/dt$, que nos dice cuánto cambia el ángulo en un segundo, así como $v = ds/dt$ describe lo rápido que una cosa se mueve, o hasta dónde se mueve en un segundo. Si el ángulo se mide en radianes, la velocidad angular ω va a ser tantos radianes por segundo. Mientras mayor es la velocidad angular, más rápido gira el objeto y más rápido cambia el ángulo. Sigamos: derivemos la velocidad angular respecto al tiempo y llamemos a $\alpha = d\omega/dt = d^\theta/dt^2$ la aceleración angular. Sería el análogo de la aceleración ordinaria.

Ahora, por supuesto, vamos a tener que relacionar la dinámica de la rotación a las leyes de la dinámica de las partículas de las cuales el objeto está hecho, de manera que debemos averiguar cómo se mueve una partícula cuando la velocidad angular es tal y tal. Para hacer esto, tomemos una cierta partícula que está localizada a una distancia r del eje y digamos que está en una cierta posición $P(x, y)$ en un instante dado, de la manera usual (figura 18-1). Si en un momento Δt más tarde el ángulo del objeto total ha girado en $\Delta\theta$, esta partícula es arrastrada con él. Está al mismo radio desde O que estaba antes, pero es llevada a Q. Lo primero que desearíamos saber es en cuánto cambia la distancia x y en cuánto cambia la distancia y. Si llamamos r a OP, el largo de PQ es $r\Delta\theta$, debido a la manera en que se definen los ángulos. El cambio de x, entonces, es simplemente la proyección de $r\Delta\theta$ en la dirección x:

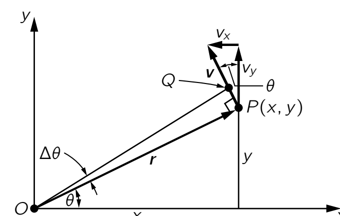


Figura 18-1. Cinemática de una rotación bidimensional.

$$\Delta x = -PQ \operatorname{sen}\theta = -r\,\Delta\theta \cdot (y/r) = -y\,\Delta\theta. \tag{18.6}$$

Análogamente,

$$\Delta y = +x\,\Delta\theta. \tag{18.7}$$

Si el objeto está girando con una velocidad angular dada ω, encontramos dividiendo ambos miembros de (18.6) y (18.7) por Δt, que la velocidad de la partícula es

$$v_x = -\omega y \qquad \text{y} \qquad v_y = +\omega x. \tag{18.8}$$

Por supuesto, si queremos encontrar el módulo de la velocidad, escribimos simplemente

$$v = \sqrt{v_x^2 + v_y^2} = \sqrt{\omega^2 y^2 + \omega^2 x^2} = \omega\sqrt{x^2 + y^2} = \omega r. \tag{18.9}$$

No debería ser un misterio que el valor del módulo de la velocidad es ωr; de hecho, debería ser patente, porque la distancia que se mueve es $r\Delta\theta$ y la distancia que se mueve por segundo es $r\Delta\theta/\Delta t$ o $r\omega$.

Pasemos a considerar la *dinámica* de la rotación. Aquí debemos introducir un nuevo concepto, *fuerza*. Investiguemos si podemos inventar algo que llamaremos torque (del latín *torquere*, torcer) que tiene la misma relación con la rotación que la fuerza tiene con el movimiento lineal. Una fuerza es la cosa que se necesita para hacer un movimiento lineal y la cosa que hace rotar a algo es una "fuerza rotatoria" o una "fuerza torcedora", es decir, un torque. Cualitativamente, un torque es una "torsión"; ¿qué es un torque cuantitativamente? Vamos a llegar a la teoría de los torques cuantitativamente estudiando el *trabajo* realizado al girar un objeto, porque una manera bonita de definir una fuerza es decir cuánto trabajo realiza cuando actúa durante un desplazamiento dado. Vamos a tratar de mantener la analogía entre cantidades lineales y angulares igualando el trabajo que hacemos cuando rotamos un poquito algo sobre el cual están actuando fuerzas al *torque* multiplicado por el *ángulo* que gira. En otras palabras, vamos a arreglar la definición de torque de manera que el teorema del trabajo tenga una analogía absoluta: la fuerza por la distancia es trabajo y el torque por el ángulo va a ser trabajo. Esto nos dice lo que es el torque. Consideren, por ejemplo, un cuerpo rígido de algún tipo con diversas fuerzas actuando en él, y un eje alrededor del cual el cuerpo gira. Concentrémonos primero en una fuerza y supongamos que esta fuerza se aplica en un cierto punto (x, y). ¿Qué trabajo se realizaría si rotáramos el objeto en un ángulo muy pequeño? Esto es fácil. El trabajo realizado es

$$\Delta W = F_x\,\Delta x + F_y\,\Delta y. \tag{18.10}$$

Necesitamos solamente sustituir las ecuaciones (18.6) y (18.7) para Δx y Δy para obtener

$$\Delta W = (xF_y - yF_x)\Delta\theta. \tag{18.11}$$

O sea, la cantidad de trabajo que hemos realizado es, de hecho, igual al ángulo en el cual hemos girado el objeto, multiplicado por una extraña combinación de fuerza y distancia. Esta "extraña combinación" es lo que llamamos torque. De manera que definiendo el cambio en el trabajo como el torque multiplicado por el ángulo, tenemos ahora la fórmula del torque en función de las fuerzas. (Evidentemente, el torque no es una idea completamente nueva independiente de la mecánica de Newton; el torque debe tener una definición determinada en términos de la fuerza.)

Cuando hay varias fuerzas actuando, el trabajo es, por supuesto, la suma de los trabajos hechos por todas las fuerzas de manera que ΔW va a ser la adición de un montón de términos, correspondientes a todas las fuerzas, *cada uno de los cuales es proporcional, sin embargo, a* $\Delta\theta$. Podemos sacar factor común $\Delta\theta$ y, por lo tanto, decir que el cambio en el trabajo es igual a la suma de todos los torques debido a todas las diversas fuerzas que están actuando, multiplicada por $\Delta\theta$. A esta suma podríamos llamarla torque total τ. Así, los torques se suman

mediante las leyes ordinarias del álgebra, pero más tarde veremos que esto sólo es así porque estamos trabajando en un plano. Es lo mismo que la cinemática unidimensional, donde las fuerzas simplemente se suman algebraicamente, pero sólo porque están todas en la misma dirección. Es más complicado en tres dimensiones. Así, para una rotación bidimensional.

$$\tau_i = x_i F_{yi} - y_i F_{xi} \tag{18.12}$$

y

$$\tau = \sum \tau_i. \tag{18.13}$$

Debe hacerse énfasis que el torque es respecto a un eje dado. Si se elige un eje diferente, de manera que todos los x_i y y_i cambian, el valor del torque cambia también en general.

Detengámonos ahora brevemente para notar que nuestra introducción anterior de torque a partir de la idea de trabajo nos da un resultado sumamente importante para un objeto en equilibrio: si todas las fuerzas sobre un objeto están equilibradas tanto para traslación como para rotación, entonces no solamente la *fuerza* resultante es cero, sino que también el total de todos los *torques* es cero, porque si un objeto está en equilibrio, *las fuerzas no realizan ningún trabajo para un desplazamiento pequeño.* Por lo tanto, ya que $\Delta W = \tau\Delta\theta = 0$, la suma de todos los torques es cero. De manera que existen dos condiciones para el equilibrio: que la suma de las fuerzas sea cero y que la suma de los torques sea cero. Demuestren que es suficiente asegurarse que la suma de los torques respecto a un eje cualquiera (en dos dimensiones) es cero.

Consideremos ahora una sola fuerza y tratemos de averiguar, geométricamente, a qué corresponde esta extraña cosa $xF_y - yF_x$, En la figura 18-2 vemos una fuerza $\boldsymbol{F}$ actuando en un punto $\boldsymbol{r}$. Cuando el objeto ha rotado en un pequeño ángulo $\Delta\theta$, el trabajo realizado, por supuesto, es la componente de la fuerza en la dirección del desplazamiento multiplicada por el desplazamiento. En otras palabras, es sólo la componente tangencial de la fuerza la que cuenta y ésta debe ser multiplicada por la distancia $r\Delta\theta$. Por lo tanto, vemos que el torque también es igual a la componente tangencial de la fuerza (perpendicular al radio) multiplicada por el radio. Esto tiene sentido de acuerdo con nuestra idea ordinaria de torque, porque si la fuerza fuera completamente radial, no daría ninguna "torsión" al cuerpo; es evidente que el efecto de torcer debe incluir sólo aquella parte de la fuerza que no tira por el centro y eso significa la componente tangencial. Además está claro que una fuerza dada es más efectiva con un brazo largo que cerca del eje. De hecho, si tomamos el caso de empujar justo sobre el eje, ¡no estamos torciendo en absoluto! De manera que tiene sentido que la cantidad de torsión o torque sea proporcional tanto a la distancia radial como a la componente tangencial de la fuerza.

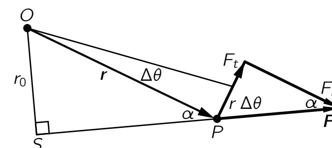


Figura 18-2. El torque producido por una fuerza.

Hay todavía una tercera fórmula para el torque que es muy interesante. Hemos visto recién que el torque es la fuerza por el radio y por el seno del ángulo α, en la figura 18-2. Pero si extendemos la línea de acción de la fuerza y dibujamos la línea OS, la distancia perpendicular a la línea de acción de la fuerza (el *brazo de palanca* de la fuerza) notamos que este brazo de palanca es más corto que r justamente en la misma proporción que la parte tangencial de la fuerza es menor que la fuerza total. Por lo tanto, la fórmula del torque también se puede escribir como el módulo de la fuerza por el largo del brazo de palanca.

El torque se llama también a menudo el *momento* de la fuerza. El origen de este término es oscuro, pero puede ser relacionado al hecho que "momento" se deriva del latín *movimentum*, y que la capacidad de una fuerza para mover un objeto (usando la fuerza en una palanca) aumenta con el largo del brazo de palanca. En matemáticas, "momento" significa ponderado por lo alejado que está del eje.

18-3 Momentum angular

Aunque hasta ahora hemos considerado solamente el caso especial de un cuerpo rígido, las propiedades de los torques y sus relaciones matemáticas son

interesantes también, aun cuando un objeto no sea rígido. En realidad podemos demostrar un teorema muy notable: de la misma manera que una fuerza externa es la rapidez de variación de una cantidad p que llamamos momentum total de un conjunto de partículas, así el torque externo es la rapidez de variación de una cantidad Ls que llamamos *momentum angular* del grupo de partículas.

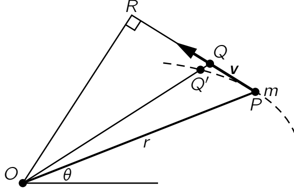


Figura 18-3. Una partícula se mueve alrededor de un eje O.

Para demostrar esto, vamos a suponer que hay un sistema de partículas en el cual hay algunas fuerzas actuando y averiguar qué sucede al sistema como consecuencia de los torques debido a estas fuerzas. Primero, por supuesto, debemos considerar *una* sola partícula. En la figura 18-3 hay una partícula de masa m y un eje O; la partícula no está rotando necesariamente en un círculo cerca de la O, puede estarse moviendo en una elipse, como un planeta alrededor del Sol, o en cualquier otra curva. Se mueve de alguna manera y hay fuerzas sobre ella y acelera de acuerdo con la fórmula usual que la componente x de la fuerza es la masa por la componente x de la aceleración, etc. Pero veamos qué hace el *torque*. El torque es igual a $xF_y - yF_x$ y la fuerza en la dirección x o y es la masa por la aceleración en la dirección x o y:

$$\begin{aligned} \tau &= xF_y - yF_x = \\ &= xm(d^2y/dt^2) - ym(d^2x/dt^2). \end{aligned} \tag{18.14}$$

Ahora, aunque esto no parece ser la derivada de ninguna cantidad simple, es en realidad la derivada de la cantidad $xm(dy/dt) - ym(dx/dt)$:

$$\begin{aligned} \frac{d}{dt}\left[xm\left(\frac{dy}{dt}\right) - ym\left(\frac{dx}{dt}\right)\right] &= xm\left(\frac{d^2y}{dt^2}\right) + \left(\frac{dx}{dt}\right)m\left(\frac{dy}{dt}\right) \\ -ym\left(\frac{d^2x}{dt^2}\right) - \left(\frac{dy}{dt}\right)m\left(\frac{dx}{dt}\right) &= xm\left(\frac{d^2y}{dt^2}\right) - ym\left(\frac{d^2x}{dt^2}\right). \end{aligned} \tag{18.15}$$

¡De manera que es cierto que el torque es la variación de algo en el tiempo! Nos fijamos entonces en el "algo", le damos un nombre: lo llamamos L, la cantidad de movimiento angular:

$$\begin{aligned} L &= xm(dy/dt) - ym(dx/dt) \\ &= xp_y - yp_x. \end{aligned} \tag{18.16}$$

Aunque esta discusión no es relativista, la segunda forma para L dada más arriba es correcta desde el punto de vista relativista. ¡Hemos encontrado así que también hay un análogo rotacional para el momentum y que este análogo, el momentum angular, está dado por una expresión en términos de las componentes del momentum lineal que es justamente igual que la fórmula para el torque en términos de las componentes de la fuerza! Así, si queremos conocer el momentum angular de una partícula respecto a un eje, tomamos solamente la componente tangencial del momentum y la multiplicamos por el radio. En otras palabras, lo que vale para el momentum angular no es cuán rápido se mueve *desde* el origen, sino cuánto se mueve *alrededor* del origen. Sólo la parte tangencial del momentum vale para el momentum angular. Además mientras más lejos está la línea del momentum mayor es el momentum angular. Y también, ya que los hechos geométricos son los mismos ya sea que la cantidad esté indicada por p o F, es cierto que hay un brazo de palanca (¡*no* el mismo brazo de palanca de la fuerza en la partícula!) que se obtiene al extender la línea del *momentum* y encontrar la distancia perpendicular al eje. Luego, el momentum angular es el módulo del momentum por el brazo de palanca del momentum. De manera que tenemos tres fórmulas para el momentum angular del mismo modo que tenemos tres fórmulas para el torque:

$$\begin{aligned} L &= xp_y - yp_x \\ &= rp_{\text{tang}} \\ &= p \cdot \text{brazo de palanca}. \end{aligned} \tag{18.17}$$

Lo mismo que el torque, el momentum angular depende de la posición del eje respecto al cual se va a calcular.

Antes de proseguir al tratamiento de más de una partícula, apliquemos los resultados anteriores a un planeta girando alrededor del Sol. ¿En qué dirección está la fuerza? La fuerza está dirigida hacia el Sol. ¿Cuál es, entonces, el torque en el objeto? Por supuesto, esto depende de donde tomemos el eje, pero obtenemos un resultado muy simple si lo tomamos en el Sol mismo, ya que el torque es la fuerza por el brazo de palanca o la componente de la fuerza perpendicular a r multiplicado por r. Pero no hay fuerza tangencial, ¡luego no hay torque respecto a un eje en el Sol! Por lo tanto, el momentum angular del planeta que gira alrededor del Sol debe quedar constante. Veamos lo que eso significa. La componente tangencial de la velocidad por la masa y por el radio va a ser constante, porque esto es el momentum angular, y la rapidez de variación del momentum angular es el torque y en este problema el torque es cero. Por supuesto, ya que la masa también es una constante, esto significa que la velocidad tangencial por el radio es una constante. Pero esto es algo que ya sabíamos en el movimiento de un planeta. Supongan que consideramos una pequeña cantidad de tiempo Δt. ¿Cuánto va a avanzar el planeta cuando se mueve de P a Q (figura 18-3)? ¿Qué *área* va a barrer? Despreciando la muy pequeña área $QQ'P$ frente al área OPQ que es mucho más grande, es simplemente la mitad de la base PQ por la altura OR. En otras palabras, el área barrida en la unidad de tiempo va a ser igual a la velocidad por el brazo de palanca de la velocidad (por un medio). Así, la velocidad de cambio del área es proporcional al momentum angular, que es constante. De manera que la ley de Kepler sobre áreas iguales en tiempos iguales es una descripción en palabras del enunciado de la ley de conservación del momentum angular, cuando no hay torque producido por la fuerza.

18-4 Conservación del momentum angular

Ahora vamos a considerar qué sucede cuando hay un gran número de partículas, cuando un objeto está hecho de muchos pedazos con muchas fuerzas actuando entre ellos y sobre ellos desde el exterior. Por supuesto, sabemos ya, que alrededor de cualquier eje fijo dado, el torque sobre la partícula i (que es la fuerza sobre la partícula i por el brazo de palanca de esa fuerza) es igual a la rapidez de variación del momentum angular de esa partícula y que el momentum angular de la partícula i es su momentum por el brazo de palanca de su momentum. Supongamos ahora que sumamos los torques τ_i para todas las partículas y lo llamamos torque total τ. Entonces éste va a ser la variación de la suma de los momenta angulares L_i de todas las partículas, y esto define una nueva cantidad que llamamos el momentum angular total L. Así como el *momentum* total de un objeto es la suma de los momenta de todas sus partes, así el momentum angular es la suma de los momenta angulares de todas las partes. Luego la variación del L total es el torque total:

$$\tau = \sum \tau_i = \sum \frac{dL_i}{dt} = \frac{dL}{dt}. \tag{18.18}$$

Ahora bien, podría parecer que el torque total es una cosa complicada. Están todas esas fuerzas internas y todas las fuerzas externas a considerar. Pero si hacemos que la ley de Newton de acción y reacción diga no simplemente que la acción y reacción son iguales, sino también que están *dirigidas en sentidos exactamente opuestos a lo largo de la misma línea* (Newton pudo haber dicho esto realmente o no, pero lo supuso tácitamente), entonces los dos torques sobre los objetos que están reaccionando, debido a sus interacciones, van a ser iguales y opuestos porque los brazos de palanca para cualquier eje son iguales. Por lo tanto, los torques internos se anulan de a pares, y así tenemos el notable teorema que *¡la rapidez de variación del momentum angular total respecto a cualquier eje es igual al torque externo respecto a ese eje!*

$$\tau = \sum \tau_i = \tau_{\text{ext}} = dL/dt. \tag{18.19}$$

Tenemos entonces un teorema muy poderoso referente al movimiento de un gran conjunto de partículas, que nos permite estudiar al movimiento en conjunto sin tener que examinar la detallada maquinaria que hay adentro. Este teorema es válido para cualquier conjunto de objetos, formen un cuerpo rígido o no.

Un caso sumamente importante del teorema anterior es la *ley de conservación del momentum angular*: si ningún torque externo actúa sobre un sistema de partículas, el momentum angular se mantiene constante.

Un caso especial de gran importancia es el de un cuerpo rígido, esto es, un objeto de una forma definida que está rotando. Considérese un objeto que tiene sus dimensiones geométricas fijas y que está rotando en torno a un eje fijo. Las diversas partes del objeto mantienen la misma relación entre sí en todo instante. Tratemos ahora de encontrar el momentum angular total de este objeto. Si la masa de una de sus partículas es m_i, y su posición o localización está en (x_i, y_i), entonces el problema es encontrar el momentum angular de esa partícula, ya que el momentum angular total es la suma de los momenta angulares de tales partículas en el cuerpo. Para un objeto que se mueve en un círculo, el momentum angular, por supuesto, es la masa por la velocidad por la distancia al eje, y la velocidad es igual a la velocidad angular por la distancia al eje:

$$L_i = m_i v_i r_i = m_i r_i^2 \omega, \tag{18.20}$$

o sumando respecto a todas las partículas i, obtenemos

$$L = I\omega, \tag{18.21}$$

donde

$$I = \sum_i m_i r_i^2. \tag{18.22}$$

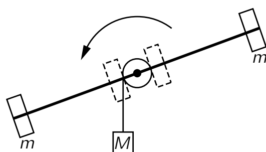


Figura 18-4. La "inercia de rotación" depende del brazo de palanca de las masas.

Éste es el análogo de la ley según la cual el momentum es la masa por la velocidad. La velocidad se remplaza por la velocidad angular y vemos que la masa queda remplazada por una cosa nueva que llamamos *momento de inercia* I, que es análogo a la masa. Las ecuaciones (18.21) y (18.22) dicen que un cuerpo tiene inercia de rotación que depende no precisamente de las masas, sino en *lo lejos que están del eje*. Así, si tenemos dos objetos con la misma masa, cuando ponemos las masas más alejadas del eje, la inercia de rotación va a ser mayor. Esto se demuestra fácilmente mediante el dispositivo mostrado en la figura 18-4, donde se impide que un peso M caiga muy rápido porque debe hacer girar una larga barra pesada. Al principio, las masas m están cerca del eje y M aumenta su velocidad a un cierto ritmo. Pero cuando cambiamos el momento de inercia al poner las masas m mucho más lejos del eje, entonces vemos que M acelera mucho menos rápidamente que antes, porque el cuerpo tiene mucho más inercia contra la rotación. El momento de inercia es la inercia contra la rotación y es la suma de las contribuciones de todas las masas por el *cuadrado* de sus distancias al eje.

Hay una diferencia importante entre masa y momento de inercia que es espectacular. La masa de un objeto no cambia nunca, pero su momento de inercia se *puede* cambiar. Si nos paramos encima de una tarima rotatoria sin roce con nuestros brazos extendidos y sostenemos unos pesos en nuestras manos mientras rotamos lentamente, podemos cambiar nuestro momento de inercia juntando nuestros brazos, pero nuestra masa no cambia. Cuando hacemos esto, todo tipo de cosas maravillosas suceden debido a la ley de conservación del momentum angular: si el torque externo es cero, entonces el momentum angular, el momento de inercia por omega, se mantiene constante. Inicialmente estábamos rotando con un momento de inercia grande I_1 a baja velocidad angular ω_1 y el momentum angular era $I_1\omega_1$. Después cambiamos nuestro momento de inercia al juntar nuestros brazos a, digamos, un valor I_2. Entonces, el producto $I\omega$, que debe mantenerse igual porque el momentum angular total debe permanecer constante, fue $I_2\omega_2$. Por lo tanto, $I_1\omega_1 = I_2\omega_2$. De manera que si *disminuimos* el momento de inercia debemos *aumentar* la velocidad angular.

19

Centro de masa; momento de inercia

19-1 Propiedades del centro de masa

En el capítulo anterior encontramos que si un gran número de fuerzas está actuando sobre una complicada masa de partículas, ya sea que las partículas integren un cuerpo rígido, o una nube de estrellas o cualquier otra cosa, y calculamos la suma de todas las fuerzas (es decir, las fuerzas externas por supuesto, porque las fuerzas internas se anulan), entonces si consideramos el cuerpo en su totalidad y suponemos que tiene una masa total M, hay un cierto punto "dentro" del cuerpo, llamado el *centro de masa* tal que la fuerza externa neta resultante produce una aceleración de este punto como si toda la masa estuviera concentrada ahí. Discutamos ahora el centro de masas con un poco de más detalle.

La ubicación del centro de masa (CM) está dada por la ecuación

$$\boldsymbol{R}_{\text{CM}} = \frac{\sum m_i \boldsymbol{r}_i}{\sum m_i}. \tag{19.1}$$

Ésta es, por supuesto, una ecuación vectorial que representa en realidad tres ecuaciones, una para cada una de las tres dimensiones. Vamos a considerar solamente la dirección x porque si la podemos entender entonces entendemos las otras dos. ¿Qué significa $X_{\text{CM}} = \sum m_i x_i / \sum m_i$? Supongamos por un momento que el objeto está dividido en pequeños pedazos, cada uno de los cuales tiene la misma masa m; entonces la masa total es simplemente el número N de pedazos por la masa de un pedazo, por ejemplo un gramo o cualquier unidad. Entonces esta ecuación dice simplemente que sumemos todas las x y después dividamos entre el número de cosas que hemos sumado: $X_{\text{CM}} = m \sum x_i / mN = \sum x_i / N$. En otras palabras, X_{CM} es el promedio de todas las x, si las masas son iguales. Pero supongamos que una de ellas sea el doble de pesada que las otras. Entonces, en esa suma, la x aparece dos veces. Esto es fácil de entender, porque podemos pensar que esta masa doble estuviera partida en dos iguales, como las otras; entonces al tomar el promedio, por supuesto, debemos contar esa x dos veces porque hay dos masas ahí. De modo que X es la posición promedio, en la dirección x de todas las masas, siendo cada masa contada un número de veces proporcional a la masa, como si estuviera dividida en "pequeños gramos". De aquí es muy fácil demostrar que X debe estar en algún lugar entre la mayor y menor x y, por lo tanto, está dentro de la envoltura que incluye todo el cuerpo. No necesita estar en el *material* del cuerpo, porque el cuerpo podría ser un círculo, como un aro, y el centro de masas está en el centro del aro y no en el aro mismo.

Por supuesto que si un cuerpo es simétrico en alguna forma, por ejemplo un rectángulo, de manera que tenga un plano de simetría, el centro de masa está en alguna parte del plano de simetría. En el caso del rectángulo hay dos planos y eso lo localiza unívocamente. Pero si es un objeto simétrico cualquiera, el centro de gravedad está en alguna parte en el eje de simetría, porque en estas circunstancias hay tantas x positivas como negativas.

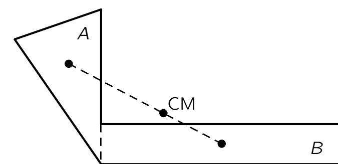


Figura 19-1. El CM de un objeto compuesto se encuentra en la línea que une los CM de las dos partes componentes.

Otra interesante proposición muy curiosa es la siguiente. Supongan que imaginemos un objeto hecho de dos pedazos A y B (figura 19-1). Entonces, el centro de masas de todo el objeto se puede calcular como sigue. Primero, encuéntrese el centro de masa del pedazo A y después el del pedazo B. Encuentre también la masa de cada pedazo M_A y M_B. Luego consideren otro problema, en el cual la masa *puntual* M_A está en el centro de masa del objeto A, y otra masa *puntual* M_B está en el centro de masa del objeto B. El centro de masa de estas

dos masas puntuales es entonces el centro de masa de todo el objeto. En otras palabras, si los centros de masa de varias partes de un objeto se han calculado, no tenemos que empezar todo de nuevo para encontrar el centro de masa del objeto entero; sencillamente tenemos que juntar los pedazos, tratando cada uno como una masa puntual situada en el centro de masa de esa pieza. Veamos por qué es así. Supongamos que queremos calcular el centro de masa de un objeto completo, algunas de cuyas partículas se consideran que forman parte de un objeto A y otras de un objeto B. La suma total $\sum m_i x_i$ puede separarse en dos partes —la suma $\sum_A m_i x_i$ para el objeto A solamente y la suma $\sum_B m_i x_i$ para el objeto B sólo—. Ahora, si estuviéramos calculando el centro de masa del objeto A sólo tendríamos exactamente la primera de estas sumas y sabemos que ésta es $M_A X_A$, la masa total de las partículas de A por la posición del centro de masa de A, porque ése es el teorema del centro de masas, aplicado al objeto A. De la misma manera, con observar el objeto B, obtenemos $M_B X_B$ y, por supuesto, sumando los dos se obtiene MX:

$$\begin{aligned} MX_{\text{CM}} &= \sum_A m_i x_i + \sum_B m_i x_i \\ &= M_A X_A + M_B X_B. \end{aligned} \tag{19.2}$$

Ahora, como M es evidentemente la suma de $M_A + M_B$, vemos que la ecuación (19.2) se puede interpretar como un ejemplo particular de la fórmula del centro de masa para dos objetos puntuales, uno de masa M_A ubicado en X_A y otro de masa M_B ubicado en X_B.

El teorema respecto al movimiento del centro de masa es muy interesante, y ha jugado una parte importante en el desarrollo de nuestra comprensión de la física. Supongan que hacemos la hipótesis que la ley de Newton es válida para las pequeñas partes componentes de un objeto mucho más grande. Entonces este teorema muestra que la ley de Newton también es correcta para un objeto más grande, aunque no estudiemos los detalles del objeto, sino solamente la fuerza total que actúa sobre él y su masa. En otras palabras, la ley de Newton tiene la propiedad peculiar que si es válida en una cierta escala pequeña también es válida en una escala mayor. Si no consideramos una pelota de béisbol como una cosa tremendamente compleja, hecha de miríadas de partículas que interactúan, sino que estudiamos sólo el movimiento del centro de masa y las fuerzas externas sobre la pelota, encontramos $\boldsymbol{F} = m\boldsymbol{a}$, donde $\boldsymbol{F}$ es la fuerza externa sobre la pelota, m su masa y $\boldsymbol{a}$ es la aceleración de su centro de masa. De manera que $\boldsymbol{F} = m\boldsymbol{a}$ es una ley que se reproduce en una escala mayor. (Debería haber una buena palabra, del griego quizás, para describir una ley que reproduzca la misma ley en una escala mayor.)

Por supuesto, uno podría sospechar que las primeras leyes que los seres humanos deberían descubrir, serían aquellas que se van a reproducir en una escala mayor. ¿Por qué? Porque la escala real de los engranajes y ruedas fundamentales del universo es de dimensiones atómicas, que son tanto más finas que nuestras observaciones, que no estamos ni cerca de esa escala en nuestras observaciones ordinarias. De manera que las primeras cosas que descubriéramos deberían ser válidas para objetos de un tamaño nada especial relativo a una escala atómica. Si las leyes para las partículas pequeñas no se reprodujeran a una escala mayor, no descubriríamos estas leyes muy fácilmente. ¿Y qué pasa con el problema inverso? ¿Deben ser las leyes en una escala pequeña las mismas que en una escala mayor? Por supuesto que no es necesariamente así en la naturaleza, que las leyes a un nivel atómico tengan que ser las mismas que en una escala mayor. Supongamos que las verdaderas leyes de movimiento de los átomos fueran dadas por alguna ecuación extraña que *no* tuviera la propiedad de que cuando vamos a una escala mayor reproduzcamos la misma ley, pero en cambio tenga la propiedad de que si vamos a una escala mayor, podamos *aproximarla mediante una cierta expresión* tal que si extendemos la expresión más y más, se sigue reproduciendo en una escala cada vez mayor. Esto es posible y de hecho es la manera en que se realiza. Las leyes de Newton son la "cola"

de las leyes atómicas, extrapoladas a un tamaño muy grande. Las leyes reales del movimiento de las partículas en una escala fina son muy peculiares, pero si tomamos un gran número de ellas y las combinamos, esas leyes se aproximan, pero *sólo* se aproximan, a las leyes de Newton. Las leyes de Newton nos permiten ir a una escala cada vez mayor y todavía parece ser la misma ley. De hecho, se hace cada vez más precisa a medida que la escala se hace mayor. Este factor autorreproductor de las leyes de Newton es así, no un aspecto fundamental de la naturaleza, sino un aspecto histórico importante. Nunca descubriríamos las leyes fundamentales de las partículas atómicas en una primera observación porque las primeras observaciones son muy toscas. De hecho, resulta que las leyes atómicas fundamentales, que llamamos mecánica cuántica, son bastante diferentes de las leyes de Newton y son difíciles de comprender porque todas nuestras experiencias directas son con objetos en una escala grande y los átomos en una escala chica se comportan de una manera que no vemos en una escala grande. De manera que no podemos decir: "un átomo es lo mismo que un planeta girando alrededor del Sol", o nada por el estilo. No se parece a *nada* con lo cual estemos familiarizados porque no hay *nada como él*. A medida que aplicamos la mecánica cuántica a objetos cada vez mayores, las leyes del comportamiento de muchos átomos juntos *no* se reproducen, pero producen *nuevas leyes*, que son las leyes de Newton, que entonces continúan reproduciéndose desde digamos tamaño de micro-micro gramo, que ya son miles de millones de átomos, hasta el tamaño de la Tierra y mayores.

Volvamos al centro de masa. El centro de masa se llama a veces centro de gravedad, porque en muchos casos la gravedad se puede considerar uniforme. Supongan que tenemos dimensiones suficientemente chicas de manera que la fuerza gravitacional no sólo es proporcional a la masa, sino además es paralela en todas partes a una línea fija. Luego consideren un objeto en el cual hay fuerzas gravitacionales en cada una de sus masas constituyentes. Sea m_i la masa de una parte. Entonces, la fuerza gravitacional en esa parte es m_i por g. La pregunta ahora es ¿dónde podemos aplicar una fuerza única para contrarrestar la fuerza gravitacional sobre el conjunto, de manera que todo el objeto, si es un cuerpo rígido, no gire? La respuesta es que esta fuerza debe pasar por el centro de masa y demostramos esto de la siguiente manera. Para que el cuerpo no gire, el torque producido por todas las fuerzas debe sumar cero, porque si hay un torque hay una variación en el momentum angular y por lo tanto una rotación. Así que debemos calcular el total de todos los torques relativo a todas las partículas y ver cuánto torque hay respecto a un eje dado; debe ser cero si este eje está en el centro de masa. Ahora, midiendo x horizontalmente y y verticalmente, sabemos que los torques son las fuerzas en la dirección y por el brazo de palanca x (es decir, la fuerza por el brazo de palanca alrededor del cual queremos medir el torque). El torque es la suma

$$\tau = \sum m_i g x_i = g \sum m_i x_i, \tag{19.3}$$

de manera que si el torque total debe ser cero, la suma $\sum m_i x_i$ debe ser cero. Pero $\sum m_i x_i = M X_{\text{CM}}$, la masa total por la distancia del centro de masa al eje. De manera que la distancia x del centro de masa desde el eje es cero.

Por supuesto, hemos comprobado el resultado sólo para la distancia x, pero si usamos el verdadero centro de masa, el objeto se va a equilibrar en cualquier posición, porque si lo giramos en 90°, tendremos y en vez de x. En otras palabras, cuando un objeto está soportado en su centro de masa, no hay torque en él debido a un campo gravitacional paralelo. En el caso de que el objeto sea tan grande que el no paralelismo de las fuerzas gravitacionales sea apreciable, el centro en el cual se debe aplicar la fuerza equilibrante no es fácil de describir y se separa ligeramente del centro de masa. Ésta es la razón por la cual hay que distinguir entre centro de masa y centro de gravedad. El hecho de que un objeto soportado exactamente en el centro de masa esté en equilibrio en todas las posiciones tiene otra consecuencia interesante. Si en vez de la gravitación tenemos una pseudofuerza debida a la aceleración, podemos usar exactamente

el mismo procedimiento matemático para encontrar la posición donde sujetarlo de manera que no haya torque producido por la fuerza inercial de la aceleración. Supongan que el objeto se mantiene de alguna manera dentro de una caja y que la caja y todo lo que contiene está acelerando. Sabemos que, desde el punto de vista de alguien en reposo relativo a esta caja acelerada, habrá una fuerza efectiva debida a la inercia. Esto es, para lograr que el objeto siga con la caja debemos empujarlo para acelerarlo y esta fuerza es "equilibrada" por la "fuerza de inercia", que es una pseudofuerza igual a la masa por la aceleración de la caja. Para el hombre en la caja, esto es lo mismo que si el objeto estuviera en un campo gravitacional uniforme cuyo valor de "g" es igual a la aceleración a. Por lo tanto, la fuerza inercial debida a la aceleración de un objeto no tiene torque respecto al centro de masa.

Este hecho tiene una consecuencia muy interesante. En un sistema inercial que no esté acelerando, el torque es siempre igual a la rapidez de variación del momentum angular. Sin embargo, respecto a un eje que pasa por el centro de masa de un objeto que *está* acelerando, *todavía es válido* que el torque es igual a la rapidez de variación del momentum angular. Aún si el centro de masa está acelerando podemos elegir un eje especial, a saber, el que pasa por el centro de masa de manera que siga siendo verdadero que el torque es igual a la rapidez de variación del momentum angular respecto a ese eje. De manera que el teorema que el torque es igual a la rapidez de variación del momentum angular es válido en dos casos generales: (1) un eje fijo en un espacio inercial; (2) un eje a través del centro de masa, aunque el objeto esté acelerando.

19-2 Cómo ubicar el centro de masa

La técnica matemática para el cálculo de centros de masa queda en al ámbito de un curso de matemática y estos problemas proporcionan un buen ejercicio en el cálculo integral. Sin embargo, después que uno ha aprendido el cálculo integral, y desea saber cómo localizar el centro de masa, es bueno conocer algunos trucos que pueden usarse para eso. Uno de estos trucos hace uso de lo que se llama teorema de Pappus. Funciona así: si tomamos un área cerrada cualquiera en un plano y generamos un sólido moviéndola en el espacio de manera que cada punto siempre se mueve perpendicular al plano del área, ¡el sólido resultante tiene un volumen igual al área de la sección por la distancia que el centro de masa se ha movido! Por cierto que esto es válido si movemos el área en una línea recta perpendicular a sí misma, pero si la movemos en un círculo o en cualquier otra curva entonces genera un volumen bastante peculiar. En una trayectoria curva la parte de afuera gira más lejos y la parte de adentro gira más cerca y los efectos se compensan. De manera que si queremos localizar el centro de masa de una lámina plana de densidad uniforme podemos recordar que el volumen generado al hacerla rotar alrededor de un eje es igual a la distancia que gira el centro de masa por el área de la lámina.

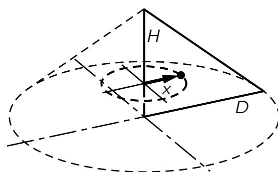


Figura 19-2. Un triángulo rectángulo y un cono circular generado por rotación del triángulo.

Por ejemplo, si queremos encontrar el centro de masa de un triángulo rectángulo de base D y altura H (figura 19-2), podemos resolver el problema de la siguiente manera. Imaginen un eje a lo largo de H y roten el triángulo alrededor de ese eje en 360 grados. Esto genera un cono. La distancia que la coordenada x del centro de masa se ha movido es $2\pi x$. El área que se ha movido es el área del triángulo $\frac{1}{2}HD$. De manera que la distancia x del centro de masa por el área del triángulo es el volumen barrido, que es por supuesto $\pi D^2 H/3$. Así $(2\pi x)(\frac{1}{2}HD) = \pi D^2 H/3$, o $x = D/3$. De una manera análoga, rotando alrededor del otro eje o por simetría encontramos $y = H/3$. De hecho, el centro de masa de cualquier área uniforme triangular está donde se cortan las medianas, las líneas que van desde los vértices hasta la mitad de los lados opuestos. Ese punto está a 1/3 de cada mediana. *Clave*: rebanen el triángulo en pequeños pedazos paralelos a una base. Noten que la mediana bisecta cada pedazo y, por lo tanto, el centro de masa debe estar en esta línea.

Tratemos ahora una figura más complicada. Supongamos que se desea encontrar la posición del centro de masa de un disco semicircular uniforme —un

disco partido por la mitad—. ¿Dónde está el centro de masa? Para un disco completo está en el centro, por supuesto, pero para medio disco es más difícil. Sea r el radio y x la distancia del centro de masa desde el borde recto del disco. Gírenlo alrededor de este borde como un eje para generar una esfera. Entonces, el centro de masa ha girado en $2\pi x$; el área es $\pi r^2/2$ (porque es sólo la mitad de un círculo). El volumen generado es, por supuesto, $4\pi r^3/3$, de donde encontramos

$$(2\pi x)(\tfrac{1}{2}\pi r^2) = 4\pi r^3/3,$$

o

$$x = 4r/3\pi.$$

Hay otro teorema de Pappus que es un caso especial del anterior y, por lo tanto, igualmente válido. Supongamos que, en vez de un disco sólido semicircular, tengamos un pedazo de alambre semicircular con densidad de masa uniforme a lo largo del alambre, y queremos encontrar su centro de masa. En este caso no hay masa en el interior, solamente en el alambre. Entonces resulta que el *área* barrida por una curva plana, cuando se mueve como antes, es la distancia que el centro de masa se mueve por el *largo* de la línea. (La línea se puede considerar como un área muy angosta y el teorema anterior se le puede aplicar.)

19-3 Cómo obtener el momento de inercia

Discutamos ahora el problema de encontrar los *momentos de inercia* de varios objetos. La fórmula para el momento de inercia alrededor del eje z de un objeto es

$$I = \sum m_i(x_i^2 + y_i^2)$$

o

$$I = \int (x^2 + y^2)\,dm = \int (x^2 + y^2)\rho\,dV. \tag{19.4}$$

O sea, debemos sumar las masas cada una multiplicada por el cuadrado de su distancia $(x_i^2 + y_i^2)$ al eje. Noten que no es la distancia tridimensional, solamente la distancia bidimensional al cuadrado aun para un objeto tridimensional. En la mayor parte nos vamos a restringir a objetos bidimensionales, pero la fórmula para la rotación alrededor del eje z es la misma en tres dimensiones.

Como un ejemplo simple, consideren una barra que rota alrededor de un eje perpendicular a uno de sus extremos (figura 19-3). Ahora debemos sumar todas las masas multiplicadas por las distancias x al cuadrado (siendo todas las y cero en este caso). Por supuesto, lo que entendemos por "suma" es la integral de x^2 por los pequeños elementos de masa. Si dividimos la barra en pequeños elementos de largo dx, los elementos de masas correspondientes son proporcionales a dx, y si dx fuera el largo de toda la barra, la masa sería M. Por lo tanto,

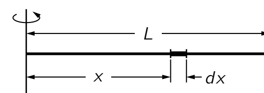


Figura 19-3. Una barra recta de largo L rotando con respecto a un eje que pasa por un extremo.

$$dm = M\,dx/L$$

y así

$$I = \int_0^L x^2\,\frac{M\,dx}{L} = \frac{M}{L}\int_0^L x^2\,dx = \frac{ML^2}{3}. \tag{19.5}$$

Las dimensiones del momento de inercia son siempre masa por distancia al cuadrado, así que todo lo que tuvimos que encontrar fue el factor 1/3.

Ahora bien, ¿cuánto vale I si el eje de rotación está en el centro de la barra? Podríamos simplemente realizar la integral de nuevo haciendo variar x de $-\frac{1}{2}L$ a $+\frac{1}{2}L$. Pero notemos algunas cosas acerca del momento de inercia. Podemos imaginar la barra como dos barras cada una de masa $M/2$ y largo $L/2$; los momentos de inercia de las dos barras pequeñas son iguales y ambos están dados por la fórmula (19.5). Luego el momento de inercia es

$$I = \frac{2(M/2)(L/2)^2}{3} = \frac{ML^2}{12}. \tag{19.6}$$

Por lo tanto, es más fácil girar una barra alrededor de su centro que balancearla alrededor de un extremo.

Podríamos, por supuesto, proseguir calculando los momentos de inercia de varios otros cuerpos de interés. Sin embargo, aunque estos cálculos proporcionan una cierta cantidad importante de ejercicios en el cálculo integral, no son básicamente de interés como tales para nosotros. Sin embargo, hay un teorema que es muy útil. Supongan que tenemos un objeto y queremos encontrar su momento de inercia respecto a algún eje. Eso significa que queremos la inercia necesaria para ponerlo en rotación alrededor de ese eje. Si sujetamos el objeto en pivotes en el centro de masa de manera que el objeto no gire a medida que rota alrededor del eje (porque no hay torque en él debido a efectos inerciales y por lo tanto no va a girar cuando empezamos a mover), entonces las fuerzas necesarias para rotarlo son las mismas que si la masa estuviera concentrada en el centro de masa y el momento de inercia sería simplemente $I_1 = MR_{\text{CM}}^2$, donde R_{CM} es la distancia desde el eje al centro de masa. Pero, por supuesto, ésa no es la fórmula correcta para el momento de inercia de un objeto que realmente está rotando a medida que da vuelta, porque no solamente está el centro de masa moviéndose en un círculo que contribuirá una cantidad I_1 al momento de inercia, sino que también tenemos que girarlo respecto a su centro de masa. Así que no es irracional que debamos sumar a I_1 el momento de inercia I_c alrededor del centro de masa. Así que es una buena suposición que el momento de inercia total alrededor de cualquier eje será

$$I = I_c + MR_{\text{CM}}^2. \tag{19.7}$$

Este teorema se llama *teorema de los ejes paralelos* y se puede demostrar fácilmente. El momento de inercia respecto a cualquier eje es la suma de la masa por la suma de las x_i y las y_i cada una al cuadrado: $I = \sum(x_i^2 + y_i^2)m_i$. Vamos a concentrarnos sobre las x; pero, por supuesto, las y funcionan igual. Ahora, x es la distancia de una cierta masa puntual desde el origen, pero veamos cómo aparecería si midiéramos x' desde el CM en vez de x desde el origen. Para estar listos para este análisis escribimos

$$x_i = x_i' + X_{\text{CM}}.$$

Entonces simplemente elevamos al cuadrado

$$x_i^2 = x_i'^2 + 2X_{\text{CM}}x_i' + X_{\text{CM}}^2.$$

Así, ¿qué sucede cuando se multiplica esto por m_i y se suma para todo i? Sacando las constantes afuera del signo de suma obtenemos

$$I_x = \sum m_i x_i'^2 + 2X_{\text{CM}} \sum m_i x_i' + X_{\text{CM}}^2 \sum m_i.$$

La tercera suma es fácil; es simplemente MX_{CM}^2. En la segunda suma hay dos partes, una de ellas es $\sum m_i x_i'$ que es la masa total por la coordenada x' del centro de masa. Pero esto no contribuye nada porque x' está medida desde el centro de masa y en estos ejes la posición promedio de todas las partículas ponderadas por las masas es cero. La primera suma es, por supuesto, la parte x de I_c. Así llegamos a la ecuación (19.7) como habíamos adivinado.

Verifiquemos (19.7) como un ejemplo. Veamos si resulta para la barra. Para un eje que pasa por un extremo, el momento de inercia debe ser $ML^2/3$ porque así lo calculamos. El centro de masa de una barra está, por supuesto, en el centro de la barra a una distancia $L/2$. Por lo tanto, deberíamos encontrar que $ML^2/3 = ML^2/12 + M(L/2)^2$. Ya que un cuarto más un doceavo es un tercio, no hemos cometido ningún error fundamental.

Entre paréntesis, en realidad no necesitábamos usar una integral para encontrar el momento de inercia (19.5). Si sencillamente suponemos que es ML^2 por γ, un coeficiente desconocido, y luego usamos el argumento acerca de las dos mitades para obtener $\frac{1}{4}\gamma$, para (19.6), entonces con nuestro razonamiento acerca de la

transferencia de ejes podríamos probar que $\gamma = \frac{1}{4}\gamma + \frac{1}{4}$, de manera que γ debe ser 1/3. ¡Siempre habrá alguna otra manera de hacerlo!

Al aplicar el teorema de los ejes paralelos es por cierto importante recordar que el eje para I_c *debe ser paralelo* al eje respecto al cual se quiere el momento de inercia.

Vale la pena mencionar otra propiedad del momento de inercia porque a menudo es útil para encontrar el momento de inercia de cierto tipo de objetos. Esta propiedad consiste en que si uno tiene una *figura plana* y un conjunto de ejes coordenados con origen en el plano y el eje z perpendicular al plano, entonces el momento de inercia de esta figura respecto al eje z es igual a la suma de los momentos de inercia respecto a los ejes x y y. Esto se demuestra fácilmente notando que

$$I_x = \sum m_i(y_i^2 + z_i^2) = \sum m_i y_i^2$$

(ya que $z_i = 0$). Análogamente,

$$I_y = \sum m_i(x_i^2 + z_i^2) = \sum m_i x_i^2,$$

pero

$$\begin{aligned} I_z = \sum m_i(x_i^2 + y_i^2) &= \sum m_i x_i^2 + \sum m_i y_i^2 \\ &= I_x + I_y. \end{aligned}$$

Como un ejemplo, el momento de inercia de una placa rectangular uniforme de masa M, ancho w y largo L respecto a un eje perpendicular a la lámina y que pasa por su centro es simplemente

$$I = M(w^2 + L^2)/12,$$

porque su momento de inercia respecto a un eje en su plano y paralelo a su largo es $Mw^2/12$, es decir, precisamente como para una barra de largo w y el momento de inercia respecto al otro eje en su plano es $ML^2/12$ precisamente como para una barra de largo L.

Para resumir, el momento de inercia de un objeto respecto a un eje dado y que llamaremos eje z tiene las siguientes propiedades:

(1) El momento de inercia es

$$I_z = \sum_i m_i(x_i^2 + y_i^2) = \int (x^2 + y^2)\, dm.$$

(2) Si un objeto está hecho de un cierto número de partes, cuyos momentos de inercia son conocidos, el momento de inercia total es la suma de los momentos de inercia de las partes.

(3) El momento de inercia respecto a cualquier eje dado es igual al momento de inercia respecto a un eje paralelo a través del CM más la masa total por el cuadrado de la distancia del eje al CM.

(4) Si el objeto es una figura plana, el momento de inercia respecto a un eje perpendicular al plano es igual a la suma de los momentos de inercia respecto a dos ejes perpendiculares entre sí que están en el plano y se cortan en el eje perpendicular.

Los momentos de inercia de unas cuantas formas elementales con densidad masa uniforme se dan en la tabla 19-1, y los momentos de inercia de algunos otros objetos que pueden deducirse de la tabla 19-1, usando las propiedades anteriores, están dados en la tabla 19-2.

Tabla 19-1

Objeto	Eje z	I_z
Barra delgada, largo L	$\perp$ a la barra en el centro	$ML^2/12$
Anillo circular delgado, radios r_1 y r_2	$\perp$ al anillo en el centro	$M(r_1^2 + r_2^2)/2$
Esfera, radios r	a través del centro	$2Mr^2/5$

Tabla 19-2

Objeto	Eje z	I_z
Lámina rect., lados a, b	$\parallel b$ en el centro	$Ma^2/12$
Lámina rect., lados a, b	$\perp$ a la lámina en el centro	$M(a^2 + b^2)/12$
Anillo delgado, radios r_1, r_2	cualquier diámetro	$M(r_1^2 + r_2^2)/4$
Paralelepípedo rect., lados a, b, c	$\parallel c$, a través del centro	$M(a^2 + b^2)/12$
Cilindro circular recto, radio r, largo L	$\parallel L$, a través del centro	$Mr^2/2$
Cilindro circular recto, radio r, largo L	$\perp L$, a través del centro	$M(r^2/4 + L^2/12)$

19-4 Energía cinética de rotación

Prosigamos ahora discutiendo la dinámica. En la analogía entre movimiento lineal y movimiento angular que discutimos en el capítulo 18 usamos el teorema del trabajo, pero no hablamos de energía cinética. ¿Cuál es la energía cinética de un cuerpo rígido que rota alrededor de un cierto eje con velocidad angular ω? No podemos adivinar inmediatamente la respuesta correcta usando nuestras analogías. El momento de inercia corresponde a la masa, la velocidad angular corresponde a la velocidad, así que la energía cinética debería ser $\frac{1}{2}I\omega^2$, y realmente lo es, como se va a demostrar ahora. Supongan que el objeto está rotando alrededor de cierto eje de manera que cada punto tiene una velocidad cuyo módulo es ωr_i, donde r_i es el radio desde el punto en particular al eje. Entonces si m_i es la masa de ese punto, la energía cinética total del todo es simplemente la suma de las energías cinéticas de todos los pequeños pedazos:

$$T = \tfrac{1}{2}\sum m_i v_i^2 = \tfrac{1}{2}\sum m_i (r_i\omega)^2.$$

Ahora bien, ω^2 es una constante, la misma para todos los puntos. Así

$$T = \tfrac{1}{2}\omega^2 \sum m_i r_i^2 = \tfrac{1}{2}I\omega^2. \tag{19.8}$$

Al final del capítulo 18 indicamos que existen algunos fenómenos interesantes asociados con un objeto que no es rígido, pero que cambia de una configuración rígida con un momento de inercia definido a otra configuración rígida. A saber, en nuestro ejemplo de la mesa giratoria, teníamos un cierto momento de inercia I_1, con nuestros brazos extendidos y una cierta velocidad angular ω_1. Cuando acercamos los brazos, teníamos un momento de inercia I_2 y una velocidad angular diferente ω_2, pero de nuevo estábamos "rígidos". El momentum angular se mantenía constante, ya que no había torque respecto al eje vertical de la mesa giratoria. Esto significa que $I_1\omega_1 = I_2\omega_2$. ¿Y qué pasa con la energía? Ésta es una pregunta interesante. Con nuestros brazos recogidos giramos más rápido, pero nuestro momento de inercia es menor y parece como si las energías pudieran ser iguales. Pero no lo son, porque lo que se contrapesa es I_ω y no $I_{\omega 2}$. Así, si comparamos la energía cinética antes y después, la energía cinética antes es $\frac{1}{2}I_1\omega_1^2 = \frac{1}{2}L\omega_1$, donde $L = I_1\omega_1 = I_2\omega_2$ es el momentum angular. Después,

mediante el mismo razonamiento, tenemos $T = \frac{1}{2}L\omega_2$ y como $\omega_2 > \omega_1$, la energía cinética de rotación es mayor de lo que era antes. De manera que teníamos una cierta energía cuando nuestros brazos estaban extendidos y cuando los recogíamos estábamos girando más rápido y teníamos más energía cinética. ¿Qué pasó con el teorema de conservación de la energía? Alguien debe haber realizado un trabajo. ¡Nosotros realizamos trabajo! ¿Cuándo realizamos ese trabajo? Cuando movemos un peso horizontalmente, no realizamos ningún trabajo. Si sujetamos algo y lo acercamos no realizamos trabajo. ¡Pero eso es cuando no estamos rotando! Cuando *estamos* rotando hay una fuerza centífruga en los pesos. Están tratando de alejarse, de manera que cuando estamos girando tenemos que atraer los pesos con resistencia de la fuerza centrífuga. De manera que el trabajo que realizamos en contra de la fuerza centrífuga debería estar de acuerdo con la diferencia en la energía de rotación y por supuesto lo está. De ahí viene la energía cinética adicional.

Hay todavía otro aspecto interesante que podemos tratar sólo descriptivamente como una cosa de interés general. Este aspecto es un poco más avanzado, pero vale la pena indicarlo porque es bastante curioso y produce efectos interesantes.

Consideren el experimento de la mesa rotatoria de nuevo. Consideren el cuerpo y los brazos separadamente, desde el punto de vista del hombre que está rotando. Después que los pesos se han acercado, todo el objeto está rotando más rápidamente; pero, observen, *la parte central del cuerpo no ha cambiado*; sin embargo, está rotando más rápido después del suceso que antes. De manera que si dibujamos un círculo alrededor del cuerpo interno y consideramos objetos dentro del círculo solamente *su* momentum angular va *a cambiar*; ellos van más rápido. Por lo tanto, debe existir un torque ejercido sobre el cuerpo mientras encogemos nuestros brazos. Ningún torque puede ser ejercido por la fuerza centrífuga, porque es radial. Esto significa que entre las fuerzas que se desarrollan en un sistema en rotación, la fuerza centrífuga no es toda la historia, *existe otra fuerza.* Esta otra fuerza se llama *fuerza de coriolis* y tiene la propiedad muy extraña de que cuando movemos algo en un sistema en rotación, parece que lo empujaran hacia el lado. Al igual que la fuerza centrífuga es una fuerza aparente. Pero si vivimos en un sistema que está rotando y movemos algo radialmente, encontramos que también debemos empujarlo hacia el lado para moverlo radialmente. Este empuje hacia el lado que tenemos que ejercer es lo que giró nuestro cuerpo.

Desarrollemos una fórmula para mostrar cómo funciona realmente la fuerza de Coriolis. Supongan que Pedro está sentado sobre un carrusel que a él le parece en reposo. Pero desde el punto de vista de Juan, que está parado en el suelo y que conoce las leyes correctas de la mecánica, el carrusel está dando vueltas. Supongan que hemos dibujado una línea radial en el carrusel y que Pedro está moviendo una masa radialmente a lo largo de esta línea. Desearíamos demostrar que se necesita una fuerza hacia el lado para hacerlo. Podemos hacerlo observando el momentum angular de la masa. Ella está siempre dando vueltas con la misma velocidad angular que ω, de manera que el momentum angular es

$$L = mv_{\text{tang}}r = m\omega r \cdot r = m\omega r^2.$$

Así, cuando la masa está cerca del centro, tiene un momentum angular relativamente pequeño, pero si nos movemos a una posición más lejana, si aumentamos r, m tiene mayor momentum angular de manera que *se debe ejercer un torque* para moverlo a lo largo del radio. (Para caminar a lo largo de un radio en un carrusel uno tiene que inclinarse y empujar hacia el lado. Inténtenlo alguna vez.) El torque que se necesita es la rapidez de variación de L en el tiempo a medida que m se mueve a lo largo del radio. Si m se mueve solamente a lo largo del radio, omega se mantiene constante de manera que el torque es

$$\tau = F_c r = \frac{dL}{dt} = \frac{d(m\omega r^2)}{dt} = 2m\omega r\,\frac{dr}{dt},$$

donde F_c es la fuerza de Coriolis. Lo que realmente deseamos saber es qué fuerza hacia el lado debe ejercer Pedro para que mueva m con una velocidad $v_r = dr/dt$. Es $F_c = \tau/r = 2m\omega v_r$.

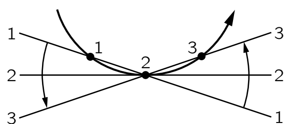


Figura 19-4. Tres vistas sucesivas de un punto que se mueve radialmente en una mesa giratoria que está rotando.

Ahora que tenemos una fórmula para la fuerza de Coriolis, examinemos esta situación con un poco más de cuidado, para ver si podemos entender el origen de esta fuerza desde un punto de vista más elemental. Observemos que la fuerza de Coriolis es la misma para cualquier radio ¡y está presente aún en el origen! Pero es especialmente fácil de entender en el origen simplemente mirando qué sucede desde el sistema inercial de Juan que está parado en el suelo. La figura 19-4 muestra tres vistas sucesivas de m justo cuando pasa por el origen para $t = 0$. Debido a la rotación del carrusel vemos que m no se mueve en línea recta, sino que en una *trayectoria curva* tangente a un diámetro del carrusel donde $r = 0$. Para que m se mueva en una curva debe existir una fuerza que la acelere en el espacio absoluto. Ésta es la fuerza de Coriolis.

Éste no es el único caso en el cual la fuerza de Coriolis aparece. También podemos demostrar que si un objeto se mueve con velocidad constante en una cierta circunferencia, también hay una fuerza de Coriolis. ¿Por qué? Pedro ve una velocidad v_M alrededor de un círculo. Por otra parte Juan ve a m moviéndose alrededor de un círculo con velocidad $v_j = v_M + \omega r$, porque m también es arrastrada por el carrusel. Por lo tanto, sabemos lo que la fuerza es realmente, es decir, la fuerza centrífuga total debida a la velocidad v_J o sea $mv_J^2 r$; ésa es la fuerza real. Desde el punto de vista de Pedro, esta fuerza centrípeta consta de tres partes. Podemos escribirla como sigue:

$$F_r = -\frac{mv_J^2}{r} = -\frac{mv_M^2}{r} - 2mv_M\omega - m\omega^2 r.$$

Ahora bien, F_r es la fuerza que vería Pedro. Tratemos de entenderlo. ¿Podría Pedro percibir el primer término? “Sí”, diría él, “aunque yo no estuviera rotando, habría una fuerza centrípeta si yo tuviera que correr alrededor del círculo con velocidad v_M”. Ésta es simplemente la fuerza centrípeta que Pedro esperaría, que no tiene nada que ver con la rotación. Además, Pedro está perfectamente consciente de que hay otra fuerza centrípeta que actuaría aun en objetos que están quietos en su carrusel. Éste es el tercer término. Pero hay además de éstos otro término, o sea, el segundo término que es de nuevo $2m\omega v$. La fuerza de Coriolis F_c era tangencial cuando la velocidad era radial y ahora es radial cuando la velocidad es tangencial. De hecho, una expresión tiene un signo negativo respecto a la otra. La fuerza está siempre en la misma dirección respecto a la velocidad, cualquiera que sea la dirección en que está la velocidad. La fuerza forma un ángulo recto con la velocidad y es de módulo $2m\omega v$.

20

Rotación en el espacio

20-1 Torques en tres dimensiones

En este capítulo vamos a discutir una de las consecuencias más notables y divertidas de la mecánica, el comportamiento de una rueda en rotación. Para hacer esto debemos ampliar la formulación matemática del movimiento rotatorio, los principios del momentum angular, el torque, etc., a un espacio tridimensional. No vamos a *usar* estas ecuaciones en toda su generalidad ni estudiar todas sus consecuencias porque esto tomaría muchos años y pronto debemos abocarnos a otros temas. En un curso introductorio podemos presentar sólo las leyes fundamentales y aplicarlas a unas pocas situaciones de interés especial.

Primero, notemos que si tenemos una rotación en tres dimensiones, ya sea de un cuerpos rígidos o de cualquier otro sistema, lo que dedujimos para dos dimensiones todavía es válido. Es decir, todavía es cierto que $xF_y - yF_x$ es el torque "en el plano xy" o el torque en "torno al eje z". También resulta que este torque es todavía igual a la rapidez de variación de $xp_y - yp_x$, porque si volvemos a la deducción de la ecuación (18.15) a partir de las leyes de Newton, vemos que no hemos tenido que suponer que el movimiento era en el plano; cuando derivamos $xp_y - yp_x$, obtenemos $xF_y - yF_x$, de manera que este teorema todavía es válido. Entonces, a la cantidad $xp_y - yp_x$ la llamamos momentum angular correspondiente al plano xy, o momentum angular respecto al eje z. Asegurado esto, podemos usar cualquier otro par de ejes y obtener otra ecuación. Por ejemplo, podemos usar el plano yz y está claro, por simetría, que si simplemente sustituimos y por x y z por y encontraríamos $yF_z - zF_y$ para el torque y $yp_z - zp_y$ sería el momentum angular asociado con el plano yz. Por supuesto que podemos tener otro plano, el zx, y para éste encontraríamos $zF_x - xF_z = d/dt(zp_x - xp_z)$.

Que estas tres ecuaciones puedan ser deducidas para el movimiento de una partícula es bastante claro. Además si sumáramos cosas como los $xp_y - yp_x$ para muchas partículas y las llamáramos momentum angular total, tendríamos tres tipos para los tres planos xy, yz y zx; y si hiciéramos lo mismo con las fuerzas, hablaríamos del torque en los planos xy, yz y zx también. Así tendríamos leyes tales que el torque externo asociado con cualquier plano es igual a la rapidez de variación del momentum angular asociado con ese plano. Esto es precisamente una generalización de lo que escribimos en dos dimensiones.

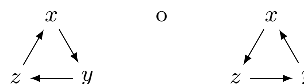
Pero alguien puede decir ahora: "Ah, pero hay más planos; después de todo, ¿no podemos acaso tomar otro plano en un cierto ángulo y calcular el torque en ese plano de las fuerzas? Ya que tendríamos que escribir otro conjunto de ecuaciones para cada uno de estos planos ¡tendríamos muchas ecuaciones!" Resulta sumamente interesante que si formáramos la combinación $x'F_{y'} - y'F_{x'}$ para otro plano, midiendo x', $F_{y'}$, etc., en ese plano, el resultado puede escribirse como una cierta *combinación* de las tres expresiones para los planos xy, yz y zx. No hay nada nuevo en ello. En otras palabras, si sabemos cuáles son los tres torques en los planos xy, yz y zx, el torque en cualquier otro plano y correspondientemente el momentum angular también se puede escribir como una combinación de éstos: 6 por 100 de uno y 92 por 100 de otro, etc. Ahora vamos a analizar esta propiedad.

Supongan que en los ejes xyz, Juan ha calculado todos sus torques y sus momenta angulares en sus planos. Pero Pedro tiene ejes x', y', z' en alguna otra dirección. Para hacerlo más sencillo, vamos a suponer que sólo los ejes x y y han sido rotados. Las x' y y' de Pedro son nuevas pero su z' es la misma. Esto es, tiene nuevos planos para, digamos, yz y zx. Por lo tanto, tiene nuevos

torques y momenta angulares para calcular. Por ejemplo, su torque en el plano $x'y'$ sería igual a $x'F_{y'} - y'F_{x'}$, etc. Lo que ahora tenemos que hacer es encontrar la relación entre los nuevos torques y los antiguos, para que podamos hacer una conexión entre un conjunto de ejes y el otro. Alguien podría decir: "Eso parece igual que lo que hicimos con vectores". Efectivamente, eso es exactamente lo que estamos intentando hacer. Entonces él puede decir: "Bueno, ¿no es el torque simplemente un vector?" *Resulta* ser un vector, pero no lo sabemos de inmediato sin hacer un análisis. Así que en las etapas siguientes vamos a hacer el análisis. No vamos a discutir cada etapa en detalle porque sólo queremos ilustrar cómo funciona. Los torques calculados por Juan son:

$$\begin{aligned}\tau_{xy} &= xF_y - yF_x,\\ \tau_{yz} &= yF_z - zF_y,\\ \tau_{zx} &= zF_x - xF_z.\end{aligned}\tag{20.1}$$

Nos apartamos del tema en este punto para hacer notar que en casos como éste no puede equivocarse de signo para alguna cantidad si las coordenadas no se manejan en forma correcta. ¿Por qué no escribir $\tau_{yz} = zF_y - yF_z$? El problema surge del hecho de que el sistema de coordenadas puede ser "derecho" o "izquierdo". Habiendo elegido (arbitrariamente) un signo, digamos, para τ_{xy}, entonces la expresión para las otras dos cantidades se puede encontrar siempre intercambiando las letras xyz en cualquiera de los dos órdenes



Pedro calcula ahora los torques en su sistema:

$$\begin{aligned}\tau_{x'y'} &= x'F_{y'} - y'F_{x'},\\ \tau_{y'z'} &= y'F_{z'} - z'F_{y'},\\ \tau_{z'x'} &= z'F_{x'} - x'F_{z'}.\end{aligned}\tag{20.2}$$

Supongamos ahora que un sistema de coordenadas se rota en un ángulo fijo θ de tal manera que los ejes z y z' sean los mismos. (Este ángulo θ no tiene nada que ver con objetos en rotación o con lo que pasa dentro del sistema de coordenadas. Es simplemente la relación entre los ejes usados por un hombre y los ejes usados por el otro y se supone que es constante.) Luego, las coordenadas de los dos sistemas están relacionadas mediante

$$\begin{aligned}x' &= x\cos\theta + y\,\mathrm{sen}\,\theta,\\ y' &= y\cos\theta - x\,\mathrm{sen}\,\theta,\\ z' &= z.\end{aligned}\tag{20.3}$$

De la misma manera, como la fuerza es un vector, se transforma en el nuevo sistema igual que x, y y z, ya que una cosa es un vector si y sólo si las diversas componentes se transforman de la misma manera que x, y y z:

$$\begin{aligned}F_{x'} &= F_x\cos\theta + F_y\,\mathrm{sen}\,\theta,\\ F_{y'} &= F_y\cos\theta - F_x\,\mathrm{sen}\,\theta,\\ F_{z'} &= F_z.\end{aligned}\tag{20.4}$$

Ahora podemos averiguar cómo se transforma el torque simplemente sustituyendo x', y' y z' por las expresiones (20.3) y $F_{x'}$, $F_{y'}$, $F_{z'}$ por las dadas por (20.4). Todas en (20.2). De manera que tenemos una fila bastante larga de términos para $\tau_{x'y'}$ y (bastante sorprendente en un principio) resulta que sale $xF_y - yF_x$, que reconocemos como el torque en el plano xy:

$$
\begin{aligned}
\tau_{x'y'} &= (x\cos\theta + y\,\mathrm{sen}\,\theta)(F_y\cos\theta - F_x\,\mathrm{sen}\,\theta)\\
&\quad - (y\cos\theta - x\,\mathrm{sen}\,\theta)(F_x\cos\theta + F_y\,\mathrm{sen}\,\theta)\\
&= xF_y(\cos^2\theta + \mathrm{sen}^2\,\theta) - yF_x(\mathrm{sen}^2\,\theta + \cos^2\theta)\\
&\quad + xF_x(-\mathrm{sen}\,\theta\cos\theta + \mathrm{sen}\,\theta\cos\theta)\\
&\quad + yF_y(\mathrm{sen}\,\theta\cos\theta - \mathrm{sen}\,\theta\cos\theta)\\
&= xF_y - yF_x = \tau_{xy}.
\end{aligned}
\tag{20.5}
$$

El resultado es claro, porque si sólo giramos nuestros ejes *en el plano*, la torsión alrededor de z en ese plano no es diferente de lo que era antes, ¡porque es el mismo plano! Lo que va a ser más interesante es la expresión para $\tau_{x'y'}$ porque este plano es nuevo. Ahora hacemos exactamente lo mismo con el plano $y'z'$ y resulta:

$$
\begin{aligned}
\tau_{y'z'} &= (y\cos\theta - x\,\mathrm{sen}\,\theta)F_z\\
&\quad - z(F_y\cos\theta - F_x\,\mathrm{sen}\,\theta)\\
&= (yF_z - zF_y)\cos\theta + (zF_x - xF_z)\,\mathrm{sen}\,\theta\\
&= \tau_{yz}\cos\theta + \tau_{zx}\,\mathrm{sen}\,\theta.
\end{aligned}
\tag{20.6}
$$

Finalmente, lo hacemos para $z'x'$:

$$
\begin{aligned}
\tau_{z'x'} &= z(F_x\cos\theta + F_y\,\mathrm{sen}\,\theta)\\
&\quad - (x\cos\theta + y\,\mathrm{sen}\,\theta)F_z\\
&= (zF_x - xF_z)\cos\theta - (yF_z - zF_y)\,\mathrm{sen}\,\theta\\
&= \tau_{zx}\cos\theta - \tau_{yz}\,\mathrm{sen}\,\theta.
\end{aligned}
\tag{20.7}
$$

Queríamos obtener una regla para encontrar los torques en nuevos ejes en función de los torques en los ejes antiguos y ahora tenemos la regla. ¿Cómo podremos recordar siempre esta regla? Si examinamos cuidadosamente (20.5), (20.6) y (20.7) vemos que hay una íntima relación entre estas ecuaciones y las ecuaciones para x, y y z. Si de alguna manera pudiéramos llamar τ_{xy} la *componente* z de algo, llamémosla la componente z de $\boldsymbol{\tau}$, entonces va a estar bien, entenderíamos (20.5) como una transformación de vectores ya que la componente z no variaría, como debe ser. De la misma manera, si asociamos con el plano yz la componente x de nuestro vector recientemente inventado y con el plano zx la componente y, estas expresiones de transformación se leerían:

$$
\begin{aligned}
\tau_{z'} &= \tau_z,\\
\tau_{x'} &= \tau_x\cos\theta + \tau_y\,\mathrm{sen}\,\theta,\\
\tau_{y'} &= \tau_y\cos\theta - \tau_x\,\mathrm{sen}\,\theta,
\end{aligned}
\tag{20.8}
$$

¡que es precisamente la regla para los vectores!

Por lo tanto, hemos demostrado que podemos identificar la combinación $xF_y - yF_x$ con lo que llamamos comúnmente la componente z de un cierto vector inventado artificialmente. Aunque el torque es una torsión en el plano y no tiene *a priori* carácter de vector, matemáticamente se comporta como vector. Este vector es perpendicular al plano de la torsión y su largo es proporcional a la intensidad de la torsión. Las tres componentes de una cantidad tal se transformarán como un vector real.

Así que representamos torques por vectores; con cada plano en el cual se supone que el torque está actuando, asociamos una línea en ángulo recto, como regla. Pero "en ángulo recto" deja el signo sin especificar. Para obtener el signo correcto debemos adoptar una regla tal que nos diga que si el torque fuera en un cierto sentido en el plano xy, el eje que queremos asociar con él tenga la dirección z "hacia arriba". O sea, alguien tiene que definir "derecha" e "izquierda" para nosotros. Suponiendo que el sistema de coordenadas es x, y, z en un sistema derecho, entonces la regla va a ser la siguiente: si pensamos en la torsión como si

estuviéramos girando un tornillo con filete derecho, la dirección del vector que vamos a asociar con esa torsión es la dirección en que avanzaría el tornillo.

¿Por qué es el torque un vector? Es un milagro de buena suerte que podamos asociar un solo eje con un plano y que, por lo tanto, podamos asociar un vector con un torque; es una propiedad especial del espacio tridimensional. En dos dimensiones el torque es un escalar común y no necesita tener una dirección asociada a él. En tres dimensiones es un vector. Si tuviéramos cuatro dimensiones, tendríamos grandes dificultades porque (si tuviéramos el tiempo, por ejemplo, como cuarta dimensión) no solamente tendríamos plano xy, yz y zx, también tendríamos planos tx, ty, y tz. Habría *seis* planos y uno no puede representar seis cantidades como un vector en cuatro dimensiones.

Vamos a estar viviendo en tres dimensiones por mucho tiempo; así que está bien notar que el tratamiento matemático anterior no dependió del hecho que x era posición y F fuerza; dependió sólo de las leyes de transformación para vectores. Por lo tanto, si en vez de x usáramos la componente x de algún otro vector, no habría ninguna diferencia. En otras palabras, si fuéramos a calcular $a_xb_y - a_yb_x$, donde $\boldsymbol{a}$ y $\boldsymbol{b}$ son vectores y lo llamáramos componente z de una nueva cantidad c, entonces estas nuevas cantidades forman un vector $\boldsymbol{c}$. Necesitamos una notación matemática para la relación entre el nuevo vector, con sus tres componentes, y los vectores $\boldsymbol{a}$ y $\boldsymbol{b}$. La notación que ha sido diseñada para esto es $\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$. Tendremos entonces, además del producto escalar común en la teoría del cálculo vectorial, un nuevo tipo de producto, llamado *producto vectorial.* Así, si $\boldsymbol{c} = \boldsymbol{a} \times \boldsymbol{b}$, esto es lo mismo que escribir:

$$\begin{aligned} c_x &= a_yb_z - a_zb_y, \\ c_y &= a_zb_x - a_xb_z, \\ c_z &= a_xb_y - a_yb_x. \end{aligned} \tag{20.9}$$

Si invertimos el orden de $\boldsymbol{a}$ y $\boldsymbol{b}$, llamando $\boldsymbol{b}$ a $\boldsymbol{a}$, y $\boldsymbol{a}$ a $\boldsymbol{b}$, tendremos el signo cambiado porque c_z sería $b_xa_y - b_ya_x$. Por lo tanto, el producto vectorial no es como la multiplicación ordinaria, donde $ab = ba$; para el producto vectorial $\boldsymbol{b} \times \boldsymbol{a} = -\boldsymbol{a} \times \boldsymbol{b}$. De esto, podemos demostrar inmediatamente que si $\boldsymbol{a} = \boldsymbol{b}$, el producto vectorial es cero. Luego $\boldsymbol{a} \times \boldsymbol{a} = 0$.

El producto vectorial es muy importante para representar los aspectos de la rotación y es importante que entendamos la relación geométrica de los tres vectores $\boldsymbol{a}$, $\boldsymbol{b}$ y $\boldsymbol{c}$. Por supuesto, que la relación entre las componentes está dada por la ecuación (20.9) y de ahí uno puede determinar cuál es la relación geométrica. La respuesta es, primero, que el vector $\boldsymbol{c}$ es perpendicular tanto a $\boldsymbol{a}$ como a $\boldsymbol{b}$. (Traten de calcular $\boldsymbol{c} \cdot \boldsymbol{a}$ y vean si no se les reduce a cero.) Segundo, el módulo de $\boldsymbol{c}$ resulta ser el módulo de $\boldsymbol{a}$ por el módulo de $\boldsymbol{b}$ por el seno del ángulo entre ellos. ¿En qué sentido apunta $\boldsymbol{c}$? Imaginen que giremos $\boldsymbol{a}$ hasta $\boldsymbol{b}$ en un ángulo menor que 180°; un tornillo con filete derecho girando de esta manera avanzará en el sentido de $\boldsymbol{c}$. El hecho que digamos un tornillo "*derecho*" en vez de un tornillo "*izquierdo*" es una convención y debe recordarse siempre que si $\boldsymbol{a}$ y $\boldsymbol{b}$ son vectores "honestos" en el sentido ordinario, el nuevo tipo de "vector" que hemos creado haciendo $\boldsymbol{a} \times \boldsymbol{b}$ es artificial o de carácter ligeramente diferente de $\boldsymbol{a}$ y $\boldsymbol{b}$, pues fue formado mediante esa regla especial. Si $\boldsymbol{a}$ y $\boldsymbol{b}$ se llaman vectores ordinarios, tenemos un nombre especial para ellos, los llamamos *vectores polares.* Ejemplos de tales vectores son la coordenada $\boldsymbol{r}$, la fuerza $\boldsymbol{F}$, el momentum $\boldsymbol{p}$, la velocidad $\boldsymbol{v}$, el campo eléctrico $\boldsymbol{E}$, etc.; éstos son vectores polares ordinarios. Los vectores que incluyen un producto vectorial en su definición se llaman *vectores axiales* o *pseudovectores.* Ejemplo de pseudovectores son, por supuesto, el torque $\boldsymbol{\tau}$ y el momentum angular $\boldsymbol{L}$. Resulta que la velocidad angular $\boldsymbol{\omega}$ es un pseudovector como lo es el campo magnético $\boldsymbol{B}$.

Para completar las propiedades matemáticas de los vectores, debemos saber todas las reglas para su multiplicación, usando productos escalar y vectorial. Por ahora, vamos a necesitar muy poco de esto en nuestras aplicaciones, pero para que quede completo vamos a escribir todas las reglas de la multiplicación de vectores para que podamos usar los resultados más tarde. Éstos son:

$$
\begin{aligned}
&\text{(a)} \qquad \boldsymbol{a}\times(\boldsymbol{b}+\boldsymbol{c}) = \boldsymbol{a}\times\boldsymbol{b}+\boldsymbol{a}\times\boldsymbol{c},\\
&\text{(b)} \qquad (\alpha\boldsymbol{a})\times\boldsymbol{b} = \alpha(\boldsymbol{a}\times\boldsymbol{b}),\\
&\text{(c)} \qquad \boldsymbol{a}\cdot(\boldsymbol{b}\times\boldsymbol{c}) = (\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c},\\
&\text{(d)} \qquad \boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c}) = \boldsymbol{b}(\boldsymbol{a}\cdot\boldsymbol{c})-\boldsymbol{c}(\boldsymbol{a}\cdot\boldsymbol{b}),\\
&\text{(e)} \qquad \boldsymbol{a}\times\boldsymbol{a} = 0,\\
&\text{(f)} \qquad \boldsymbol{a}\cdot(\boldsymbol{a}\times\boldsymbol{b}) = 0.
\end{aligned}
\tag{20.10}
$$

20-2 Las ecuaciones de rotación usando productos vectoriales

Preguntémonos ahora si hay alguna ecuación en la física que pueda ser escrita usando el producto vectorial. La respuesta es, por cierto, que hay una gran cantidad de ecuaciones que se pueden escribir así. Por ejemplo, vemos inmediatamente que el torque es igual al producto vectorial de la posición por la fuerza:

$$\boldsymbol{\tau} = \boldsymbol{r}\times\boldsymbol{F}. \tag{20.11}$$

Éste es un resumen vectorial de las tres ecuaciones $\tau_x = yF_z - zF_y$, etc. Por la misma razón, el vector momentum angular, si hay una sola partícula presente, es la distancia desde el origen multiplicada por el vector momentum:

$$\boldsymbol{L} = \boldsymbol{r}\times\boldsymbol{p}. \tag{20.12}$$

Para una rotación en un espacio tridimensional, la ley dinámica análoga a la ley $\boldsymbol{F} = d\boldsymbol{p}/dt$ de Newton es que el vector torque es la derivada respecto al tiempo del vector momentum angular:

$$\boldsymbol{\tau} = d\boldsymbol{L}/dt. \tag{20.13}$$

Si sumamos (20.13) para muchas partículas, el torque externo en un sistema es la derivada del momentum angular total respecto al tiempo:

$$\boldsymbol{\tau}_{\text{ext}} = d\boldsymbol{L}_{\text{tot}}/dt. \tag{20.14}$$

Otro teorema: si el torque total externo es cero, el vector momentum angular total del sistema es una constante. Esto se llama la ley de *conservación del momentum angular*. Si no existe torque en un sistema dado, su momentum angular no puede cambiar.

¿Qué pasa con la velocidad angular? ¿Es un vector? Ya hemos discutido la rotación de un objeto sólido alrededor de un eje fijo, pero por el momento supongan que estamos girando simultáneamente respecto a *dos* ejes. Podría estar girando respecto de una caja mientras la caja está girando respecto a algún otro eje. ¡El resultado neto de estos movimientos combinados es que el objeto simplemente gira respecto a algún eje nuevo! Lo maravilloso de este nuevo eje es que se puede imaginar de la siguiente manera. Si escribimos la velocidad de rotación en el plano xy como un vector en la dirección z cuyo largo es igual a la velocidad de rotación en el plano, y si otro vector se dibuja en la dirección y, digamos, que es la velocidad de rotación en el plano zx, entonces, si los sumamos vectorialmente mediante la regla del paralelogramo, el módulo del resultado nos dice con qué velocidad está girando el objeto y la dirección nos dice en qué plano. Es decir, simplemente, la velocidad angular es un vector del cual obtenemos las magnitudes de las rotaciones en tres planos como proyecciones en ángulo recto sobre estos planos.*

Como una aplicación simple del uso del vector velocidad angular podemos calcular la potencia gastada por el torque que actúa sobre un cuerpo rígido. La

* Que esto es cierto puede deducirse componiendo los desplazamientos de las partículas del cuerpo durante un tiempo infinitesimal Δt. Esto no es evidente y se deja para aquellos que estén interesados en resolverlo.

potencia es, por cierto, la variación del trabajo con relación al tiempo; en tres dimensiones, la potencia resulta ser $P = \boldsymbol{\tau} \cdot \boldsymbol{\omega}$.

Todas las fórmulas que escribimos para una rotación plana pueden ser generalizadas a tres dimensiones. Por ejemplo, si un cuerpo rígido está girando respecto a un cierto eje con velocidad angular $\boldsymbol{\omega}$, nos podemos preguntar: "¿Cuál es la velocidad de un punto a una cierta posición radial $\boldsymbol{r}$?" Dejaremos como problema para el estudiante el demostrar que la velocidad de una partícula en un cuerpo rígido está dada por $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$, donde $\boldsymbol{\omega}$ es la velocidad angular y $\boldsymbol{r}$ la posición. También, como otro ejemplo de productos vectoriales, tuvimos una fórmula para la fuerza de coriolis, que también puede ser escrita usando productos vectoriales: $\boldsymbol{F}_c = 2m\boldsymbol{v} \times \boldsymbol{\omega}$. Es decir, si una partícula se mueve con velocidad $\boldsymbol{v}$ en un sistema de coordenadas que está en realidad rotando con velocidad angular $\boldsymbol{\omega}$ y queremos pensar en términos de un sistema coordenado en rotación, entonces debemos agregar la pseudofuerza $\boldsymbol{F}_c$.

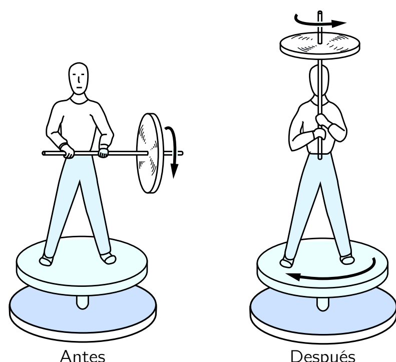


Figura 20-1. *Antes*: el eje es horizontal; momentum angular respecto al eje vertical = 0. *Después*: el eje es vertical; momentum angular respecto al eje vertical es todavía cero; el hombre y la silla giran en dirección opuesta al giro de la rueda.

20-3 El giroscopio

Volvamos ahora a la ley de conservación del momentum angular. Esta ley puede demostrarse con una rueda que gira rápidamente, o giroscopio, como sigue (véase figura 20-1). Si nos sentamos en una silla giratoria y sujetamos la rueda que gira con su eje horizontal, la rueda tiene un momentum angular respecto al eje horizontal. El momentum angular respecto a un eje *vertical* no puede cambiar debido al pivote (sin fricción) de la silla, de manera que si giramos el eje de la rueda hacia la vertical, entonces la rueda tendría momentum angular respecto al eje vertical, porque ahora está girando alrededor de este eje. Pero el *sistema* (la rueda, nosotros y la silla) *no puede tener* una componente vertical, de manera que nosotros y la silla debemos girar en la dirección opuesta a la rueda en rotación, para compensarlo.

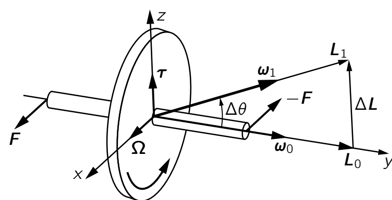


Figura 20-2. El giroscopio.

Primero analicemos con mayor detalle lo que acabamos de describir. Lo que es sorprendente y que debemos entender es el origen de las fuerzas que nos giran a nosotros y a la silla a medida que giramos el eje del giroscopio hacia la vertical. La figura 20-2 muestra la rueda girando rápidamente en torno al eje y. Luego su velocidad angular es en torno a ese eje y; resulta que su momentum angular también está en esa dirección. Supongan ahora que queremos rotar la rueda alrededor del eje x con una velocidad angular $\boldsymbol{\Omega}$; ¿qué fuerzas se necesitan? Después de un corto tiempo Δt, el eje ha girado a una nueva posición, inclinada en un ángulo $\Delta\theta$ con la horizontal. Ya que la mayor parte del momentum angular se debe a la rotación alrededor del eje (la rotación lenta contribuye muy poco), vemos que el vector momentum angular ha cambiado. ¿Cuál es el cambio del momentum angular? El momentum angular no cambia en *magnitud*, pero cambia en *dirección* una cantidad $\Delta\theta$. La magnitud del vector $\Delta\boldsymbol{L}$ es entonces $\Delta L = L_0\Delta\theta$, de manera que el torque, que es la variación en el tiempo del momentum angular, es $\tau = \Delta L/\Delta t = L_0\Delta\theta/\Delta t = L_0\Omega$. Tomando en cuenta las direcciones de las diversas cantidades vemos que

$$\boldsymbol{\tau} = \boldsymbol{\Omega} \times \boldsymbol{L}_0. \tag{20.15}$$

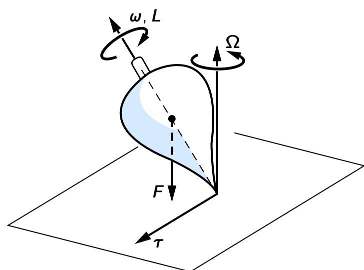


Figura 20-3. Un trompo girando rápidamente. Noten que la dirección del vector torque es la dirección de la precesión.

Por lo tanto, si $\boldsymbol{\Omega}$ y $\boldsymbol{L}_0$ son horizontales, como se muestra en la figura, $\boldsymbol{\tau}$ es *vertical*. Para producir este torque deben aplicarse a los extremos del eje unas fuerzas horizontales $\boldsymbol{F}$ y $-\boldsymbol{F}$. ¿Cómo se aplican estas fuerzas? Con nuestras manos, mientras tratamos de rotar el eje de la rueda hacia la dirección vertical. Pero la tercera ley de Newton exige que fuerzas iguales y contrarias (y *torques* iguales y opuestos) actúen en *nosotros*. Esto nos obliga a rotar en el sentido opuesto respecto al eje z vertical.

Este resultado puede ser generalizado para un trompo que gira rápidamente. En el caso familiar de un trompo que gira, la gravedad que actúa en su centro de masa proporciona un torque respecto al punto de contacto con el suelo (véase figura 20-3). Este torque es en la dirección horizontal y hace que el trompo

precese con su eje moviéndose en un cono circular respecto a la vertical. Si $\boldsymbol{\Omega}$ es la velocidad angular (vertical) de precesión, encontramos de nuevo que

$$\boldsymbol{\tau} = d\boldsymbol{L}/dt = \boldsymbol{\Omega} \times \boldsymbol{L}_0.$$

Así, cuando aplicamos un torque a un trompo que está girando muy rápidamente, la dirección del movimiento precesional es en la dirección del torque, perpendicular a las fuerzas que producen el torque.

Ahora podemos pretender que entendemos la precesión de los giroscopios y realmente lo entendemos matemáticamente. Sin embargo, ésta es una cosa matemática que en un sentido aparece como un "milagro". Va a suceder que a medida que nos adentremos en la física cada vez más avanzada, muchas cosas simples van a poder ser deducidas en forma matemática más rápidamente que lo que se las puede entender en un sentido fundamental o simple. Ésta es una característica extraña y a medida que entramos en trabajo más y más avanzado hay circunstancias en las cuales la matemática va a producir resultados que *nadie* ha sido realmente capaz de entender de una manera directa. Un ejemplo es la ecuación de Dirac que aparece en una forma muy simple y hermosa, pero cuyas consecuencias son difíciles de entender. En nuestro caso particular, la precesión de un trompo parece algo como un milagro incluyendo ángulos rectos y círculos y torsiones y tornillos derechos. Lo que debemos tratar de hacer es entenderla de una manera más física.

¿Cómo podemos explicar el torque en términos de las fuerzas reales y las aceleraciones? Notamos que cuando una rueda está precesando, las partículas que giran en la rueda no se están moviendo realmente en un plano porque la rueda está precesando (véase figura 20-4). Como explicamos anteriormente (figura 19-4), las partículas que están cruzando el eje de precesión se están moviendo en *trayectorias curvas* y esto requiere la aplicación de una fuerza lateral. Ésta es suministrada por nuestro empuje en el eje, que entonces comunica la fuerza al aro a través de los rayos. "Espere", dice alguien, "¿qué pasa con las partículas que están volviendo por el otro lado?" No toma mucho tiempo decidir que debe haber una fuerza en la *dirección opuesta* en ese lado. La fuerza neta que debemos aplicar es, por lo tanto, cero. Las *fuerzas* se anulan, pero una de ellas debe ser aplicada en un lado de la rueda y la otra debe ser aplicada en el otro lado de la rueda. Podríamos aplicar estas fuerzas directamente, pero como la rueda es sólida está permitido hacerlo empujando en el eje ya que las fuerzas pueden ser llevadas hacia arriba por los rayos.

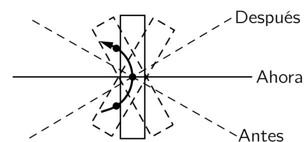


Figura 20-4. El movimiento de las partículas en la rueda en rotación de la figura 20-2, cuyo eje está rotando, es en líneas curvas.

Todo lo que hemos probado hasta ahora es que si la rueda está precesando, puede equilibrar el torque debido a la gravedad o cualquier otro torque aplicado. Pero todo lo que hemos demostrado es que esto es *una* solución de una ecuación. Esto es, si el torque está dado y si *logramos que la rotación empiece bien*, entonces la rueda va a precesar suave y uniformemente. Pero no hemos probado (y no es cierto) que una precesión uniforme es el movimiento *más general* que un cuerpo en rotación puede tener como resultado de un torque dado. El movimiento general incluye también un "tambaleo" respecto a la precesión principal. Este "tambaleo" se llama *nutación*.

A algunas personas les agrada decir que si uno ejerce un torque en un giroscopio, éste gira y precesa y que el torque *produce* la precesión. Es muy extraño que cuando uno suelta de repente un giroscopio, no *cae* bajo la acción de la gravedad ¡sino que se mueve hacia el lado! ¿Por qué sucede que la fuerza de gravedad que es hacia *abajo*, que *conocemos y sentimos*, lo hace ir hacia *un lado*? Ninguna de las fórmulas en el mundo como la (20.15) nos lo van a decir, porque (20.15) es una ecuación especial, válida solamente después que el giroscopio esté precesando muy bien. Lo que realmente sucede, en detalle, es lo siguiente. Si mantuviéramos el eje totalmente fijo, de manera que no pueda precesar de ninguna manera (pero el trompo está girando) entonces no hay ningún torque actuando, ni siquiera un torque producido por la gravedad porque está equilibrado por nuestros dedos. Pero si lo soltamos de repente, instantáneamente va a haber un torque debido a la gravedad. Cualquier persona cuerda pensaría

que el trompo va a caer, y eso es lo que empieza a hacer, lo que puede ser visto si el trompo no está girando muy rápido.

El giroscopio realmente cae, como era de esperar. Pero tan pronto como cae está girando y para que este giro continúe se necesitaría un torque. En la ausencia de un torque en esta dirección, el giroscopio empieza a "caer" en la dirección opuesta a la de la fuerza que falta. Esto da al giroscopio una componente del movimiento alrededor del eje vertical, como tendría en una precesión estable. Pero el movimiento real "sobrepasa" la velocidad de precesión estable y el eje realmente se levanta de nuevo al nivel del cual partió. La trayectoria seguida por el extremo del eje es una cicloide (la trayectoria seguida por un guijarro pegado en el neumático de un automóvil). Ordinariamente, este movimiento es muy rápido para ser seguido por el ojo y se amortigua rápidamente debido al roce en los rodamientos, dejando solamente el movimiento de precesión estable (figura 20-5). Mientras más lento gira la rueda, la nutación es más evidente.

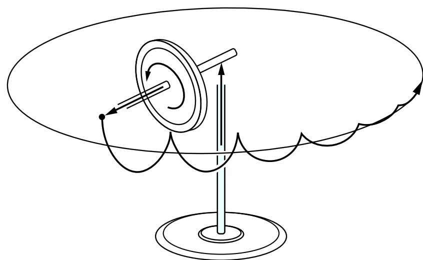


Figura 20-5. Movimiento real del extremo del eje del giroscopio bajo la acción de la gravedad inmediatamente después de soltar el eje que se había mantenido fijo previamente.

Cuando el movimiento se estabiliza, el eje del giroscopio está un poco más bajo de lo que estaba al principio. ¿Por qué? (Éstos son los detalles más complicados, pero los traemos a colación porque no queremos que el lector se forme la idea que el giroscopio es un milagro absoluto. *Es* una cosa maravillosa, pero no es un milagro.) Si estuviéramos sujetando el eje totalmente horizontal, y lo soltáramos de repente, entonces la sencilla ecuación de precesión nos diría que precesa, que gira en un plano horizontal. ¡Pero eso es imposible! Aunque lo despreciamos antes, es cierto que la rueda tiene *algún* momento de inercia respecto al eje de precesión y si se está moviendo alrededor de ese eje, aunque sea despacio, posee un débil momentum angular respecto al eje. ¿De dónde viene? Si los pivotes son perfectos no hay torque respecto al eje vertical. ¿Cómo entonces se *pone* a precesar si no hay cambio de momentum angular? La respuesta es que el movimiento cicloidal del extremo del eje se amortigua hacia el movimiento promedio estable del centro del círculo rodante equivalente. O sea, se estabiliza un poco más abajo. Porque está bajo, el momentum angular de giro tiene ahora una componente vertical pequeña, que es exactamente lo que se necesita para la precesión. Así que ven ustedes que tiene que bajar un poco para poder dar vueltas. Tiene que ceder un poco a la gravedad; al bajar su eje un poquito, mantiene la rotación alrededor del eje vertical. Ésa es, entonces, la manera en que un giroscopio funciona.

20-4 Momentum angular de un cuerpo sólido

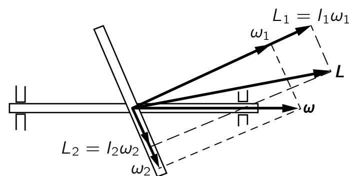


Figura 20-6. El momentum angular de un cuerpo en rotación no es necesariamente paralelo a la velocidad angular.

Antes de dejar el tema de las rotaciones en tres dimensiones, vamos a discutir, por lo menos cualitativamente, unos cuantos efectos que suceden en las rotaciones tridimensionales, los cuales no son evidentes. El efecto principal es que en general el Momentum angular de uns de un cuerpos rígidos *no está necesariamente* en la misma dirección que la velocidad angular. Consideren una rueda que está sujeta a un eje en forma asimétrica; pero, eso sí, con el eje que pasa por el centro de gravedad (figura 20-6). Cuando giramos la rueda en torno al eje, cualquiera sabe que va a haber una vibración en los rodamientos debido a la manera ladeada que la montamos. Cualitativamente, sabemos que en el sistema en rotación hay una fuerza centífruga actuando en la rueda, que trata de alejar su masa lo más posible del eje. Esto tiende a alinear el plano de la rueda de manera que sea perpendicular al eje. Para resistir esta tendencia, los rodamientos ejercen un torque. Si hay un torque ejercido por los rodamientos, debe haber una rapidez de variación del momentum angular. ¿Cómo puede haber una rapidez de variación del momentum angular cuando estamos simplemente rotando la rueda respecto al eje? Supongan que separamos la velocidad angular $\boldsymbol{\omega}$ en componentes $\boldsymbol{\omega}_1$ y $\boldsymbol{\omega}_2$ perpendicular y paralela al plano de la rueda. ¿Cuál es el momentum angular? Los momentos de inercia respecto a estos dos ejes son *diferentes*, de manera que las componentes del momentum angular, que (en estos ejes particulares y especiales solamente) son iguales a los momentos de inercia por las componentes de la velocidad angular correspondientes, están en una *razón diferente* de la que están las componentes de la velocidad angular. Por lo tanto, el vector momentum

angular está en una dirección en el espacio que *no* es a lo largo de eje. Cuando giramos el objeto, tenemos que girar el vector momentum angular en el espacio, así que debemos ejercer torques sobre el eje.

Aunque es muy complicado para demostrarlo aquí, hay una propiedad muy importante e interesante del momento de inercia que es fácil de describir y usar y que es la base de nuestro análisis anterior. Esta propiedad es la siguiente: un cuerpo rígido cualquiera, aun uno tan irregular como una papa, posee tres ejes perpendiculares entre sí a través del CM, de manera tal que el momento de inercia respecto a uno de estos ejes es el valor máximo posible para cualquier eje a través del CM, el momento de inercia respecto a otro eje tiene el *menor* valor posible y el momento de inercia respecto al tercer eje es intermedio entre estos dos (o igual a uno de ellos). Estos ejes se llaman *ejes principales* del cuerpo y tienen la importante propiedad que si el cuerpo está rotando respecto a uno de ellos, su momentum angular está en la misma dirección que su velocidad angular. Para un cuerpo con ejes de simetría, los ejes principales están según los ejes de simetría.

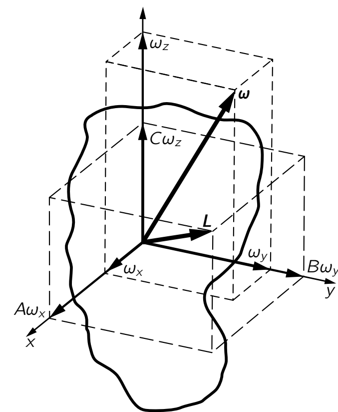


Figura 20-7. La velocidad angular y el momentum angular de un cuerpo rígido ($A > B > C$).

Si tomamos los ejes x, y y z según los ejes principales, y llamamos A, B y C a los momentos de inercia correspondientes, podemos calcular fácilmente el momentum angular y la energía cinética de rotación del cuerpo para cualquier velocidad angular $\boldsymbol{\omega}$. Si descomponemos $\boldsymbol{\omega}$ en componentes ω_x, ω_y y ω_z según los ejes x, y, z y usamos los vectores unitarios $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ también según x, y, z, podemos escribir el momentum angular como

$$\boldsymbol{L} = A\omega_x\boldsymbol{i} + B\omega_y\boldsymbol{j} + C\omega_z\boldsymbol{k}. \tag{20.16}$$

La energía cinética de rotación es:

$$\begin{aligned} \text{KE} &= \tfrac{1}{2}(A\omega_x^2 + B\omega_y^2 + C\omega_z^2) \\ &= \tfrac{1}{2}\boldsymbol{L}\cdot\boldsymbol{\omega}. \end{aligned} \tag{20.17}$$

21

El oscilador armónico

21-1 Ecuaciones diferenciales lineales

En el estudio de la física, el curso comúnmente es dividido en una serie de temas como mecánica, electricidad, óptica, etc., y uno estudia un tema después de otro. Por ejemplo, este curso hasta ahora ha tratado principalmente la mecánica. Pero una cosa rara sucede una y otra vez: las ecuaciones que aparecen en los diferentes campos de la física y aun en otras ciencias, son a menudo casi exactamente iguales, de manera que muchos fenómenos tienen analogías en estos diferentes campos. Para dar un ejemplo más sencillo, la propagación de las ondas sonoras es en muchos aspectos análoga a la propagación de las ondas luminosas. Si estudiamos acústica con gran detalle, descubrimos que mucho del trabajo es el mismo que si estuviéramos estudiando óptica con gran detalle. Así, el estudio de un fenómeno en un campo puede permitir la extensión de nuestro conocimiento en otro campo. Es mejor darse cuenta desde un principio que estas extensiones son posibles, porque de otra manera uno podría no entender la razón por la cual se gasta una gran cantidad de tiempo y energía en algo que parece ser sólo una pequeña parte de la mecánica.

El oscilador armónico, que estamos a punto de estudiar, tiene analogías íntimas en muchos otros campos; aunque empezamos con un ejemplo mecánico de una masa fija a un resorte o un péndulo con una pequeña amplitud o algunos otros dispositivos mecánicos, realmente estamos estudiando una cierta *ecuación diferencial.* Esta ecuación aparece una y otra vez en la física y en otras ciencias y de hecho pertenece a tantos fenómenos que su estudio a fondo bien vale la pena. Alguno de los fenómenos que incluye esta ecuación son las oscilaciones de una masa en un resorte; las oscilaciones de las cargas que fluyen de una parte a otra en un circuito eléctrico; las vibraciones de un diapasón que está generando ondas sonoras; las vibraciones análogas de los electrones en un átomo que generan ondas luminosas; las ecuaciones de funcionamiento de un servosistema, como un termostato tratando de ajustar una temperatura; complicadas interacciones en reacciones químicas; el crecimiento de una colonia de bacterias en interacción con el aprovisionamiento de alimento y los venenos que las bacterias producen; los zorros que se comen los conejos que se comen el pasto, etc.; todos estos fenómenos obedecen a ecuaciones que son muy similares entre sí y ésta es la razón por la cual estudiamos el oscilador mecánico con tanto detalle. Las ecuaciones se llaman *ecuaciones diferenciales lineales con coeficientes constantes.* Una ecuación diferencial lineal con coeficientes constantes es una ecuación diferencial que consiste en la suma de varios términos, siendo cada término una derivada de la variable dependiente respecto a la variable independiente y multiplicada por alguna constante. Así,

$$a_n\,d^nx/dt^n + a_{n-1}\,d^{n-1}x/dt^{n-1} + \cdots + a_1\,dx/dt + a_0x = f(t) \tag{21.1}$$

se llama una ecuación diferencial lineal de orden n con coeficientes constantes (cada a_i es constante).

21-2 El oscilador armónico

Tal vez el sistema mecánico más simple cuyo movimiento sigue una ecuación diferencial lineal con coeficientes constantes es una masa fija a un resorte: primero el resorte se estira para compensar la gravedad; una vez compensada, estudiamos

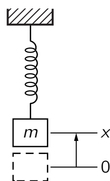


Figura 21-1. Una masa fija a un resorte: un ejemplo simple de un oscilador armónico.

el desplazamiento vertical de la masa desde su posición de equilibrio (figura 21-1). Vamos a llamar x a este desplazamiento hacia arriba y vamos a suponer también que el resorte es totalmente lineal, en cuyo caso la fuerza que tira en contra cuando el resorte está estirado es precisamente proporcional a la cantidad de estiramiento. O sea, la fuerza es $-kx$ (con un signo para recordarnos que tira en contra). Luego, la masa por la aceleración debe ser igual a $-kx$:

$$m\,d^2x/dt^2 = -kx. \tag{21.2}$$

Para simplificar, supongamos que suceda (o cambiamos nuestra unidad de medida del tiempo) que el cociente $k/m = 1$. Primero estudiaremos la ecuación

$$d^2x/dt^2 = -x. \tag{21.3}$$

Más tarde vamos a volver a la ecuación (21.2) con k y m presentes explícitamente.

Ya hemos analizado la ecuación (21.3) en detalle numéricamente; cuando recién introducimos el tema de la mecánica resolvimos esta ecuación [vean la ecuación (9.12)] para encontrar el movimiento. Mediante integración numérica encontramos una curva (figura 9-4) que mostraba que si m estaba inicialmente desplazada pero en reposo, bajaba y pasaba por cero; después no seguimos más lejos, pero por supuesto sabemos que sigue moviéndose hacia arriba y hacia abajo —*oscila*—. Cuando calculamos el movimiento numéricamente, encontramos que pasaba por el punto de equilibrio en $t = 1.570$. La duración de todo el ciclo es cuatro veces más larga; o sea $t_0 = 6.28$ "segundos". Esto fue encontrado numéricamente, antes de que supiéramos mucho cálculo integral. Suponemos que entretanto el Departamento de Matemática ha introducido una función que cuando se deriva dos veces es igual a sí misma con signo menos. (Por supuesto que hay maneras de obtener esta función de un modo directo, pero son más complicadas que saber de antemano cuál es la respuesta.) La función es $x = \cos t$. Si la derivamos encontramos $dx/dt = -\operatorname{sen} t$ y $d^2x/dt^2 = -\cos t = -x$. La función $x = \cos t$ empieza para $t = 0$ con $x = 1$ y sin velocidad inicial; ésta fue la situación con la que empezamos cuando hicimos nuestro trabajo numérico. Ahora que sabemos que $x = \cos t$, podemos calcular un valor *preciso* del tiempo en el cual debe pasar por $x = 0$. La respuesta es $t = \pi/2$ o 1.57080. Estuvimos equivocados en la última cifra debido a errores de cálculo numérico, ¡pero fue muy aproximado!

Ahora, para seguir adelante con el problema original, restauremos las unidades de tiempo a segundos reales. ¿Cuál es la solución entonces? Primero que nada, podemos pensar que podemos hacer aparecer las constantes k y m multiplicando $\cos t$ por algo. Así probamos la ecuación $x = A\cos t$, entonces encontramos $dx/dt = -A\operatorname{sen} t$, y $d^2x/dt^2 = -A\cos t = -x$. Descubrimos así con horror que no logramos resolver la ecuación (21.2) ¡sino que obtuvimos de nuevo la ecuación (21.3)! Este hecho ilustra una de las propiedades más importantes de las ecuaciones diferenciales lineales: *si multiplicamos una solución de la ecuación por una constante, esto es nuevamente una solución.* La razón matemática de esto es clara. Si x es una solución y multiplicamos ambos miembros de la ecuación por A, digamos, vemos que todas las derivadas también están multiplicadas por A y por lo tanto Ax es una solución tan buena como lo era x de la ecuación original. La física de ello es la siguiente. Si tenemos un peso fijo a un resorte y lo estiramos el doble, la fuerza también es el doble, la aceleración resultante es el doble, la velocidad que adquiere en un tiempo dado es el doble, la distancia recorrida en un tiempo dado es el doble; pero *debe* recorrer una distancia doble para volver al origen porque se estiró el doble. De manera que se demora el *mismo tiempo* en volver al origen, prescindiendo del desplazamiento inicial. En otras palabras, en una ecuación lineal el movimiento tiene el mismo *esquema temporal*, cualquiera que sea su intensidad.

Estuvo mal lo que se hizo —sólo nos enseñó que podemos multiplicar la solución por cualquier cosa y satisfacer la misma ecuación, pero no una ecuación diferente—. Después de unos pocos intentos para obtener una ecuación con una

constante diferente multiplicamos a x, encontramos que debemos alterar la escala de *tiempo*. En otras palabras, la ecuación (21.2) tiene una solución de la forma

$$x = \cos\omega_0 t. \tag{21.4}$$

(Es importante darse cuenta de que en este caso ω_0 no es una velocidad angular de un cuerpo en rotación, pero nos quedamos sin letras si no se permite usar la misma letra para más de una cosa.) La razón por la cual pusimos el subíndice "0" en ω es que vamos a tener más omegas dentro de poco; recordemos que ω_0 se refiere al movimiento natural de este oscilador. Ahora probamos la ecuación (21.4) y esta vez tenemos más éxito porque $dx/dt = -\omega_0 \operatorname{sen}\omega_0 t$ y por lo tanto $d^2x/dt^2 = -\omega_0^2 \cos\omega_0 t = -\omega_0^2 x$. Así que finalmente hemos resuelto la ecuación que realmente queríamos resolver. La ecuación $d^2x/dt^2 = -\omega_0^2 x$ es la misma que la ecuación (21.2) si $\omega_0^2 = k/m$.

Lo siguiente que tenemos que investigar es el significado físico de ω_0. Sabemos que la función coseno se repite cuando su argumento es 2π. De manera que $x = \cos\omega_0 t$ va a repetir su movimiento, va a realizar un ciclo completo, cuando el "ángulo" cambie en 2π. La cantidad $\omega_0 t$ se llama a menudo *fase* del movimiento. Para cambiar $\omega_0 t$ por 2π, el tiempo debe cambiar en una cantidad t_0 llamada el *periodo* de una oscilación completa; por supuesto, t_0 debe ser tal que $\omega_0 t_0 = 2\pi$. Esto es, $w_0 t_0$ debe corresponder a un ciclo del ángulo, entonces todo se va a repetir —si aumentamos t en t_0, agregamos 2π a la fase—. Así,

$$t_0 = 2\pi/\omega_0 = 2\pi\sqrt{m/k}. \tag{21.5}$$

De manera que si tuviéramos una masa más pesada fija a un resorte, demoraría más en oscilar de una parte a otra. Esto se debe a que tiene mayor inercia y así, aunque las fuerzas son las mismas, demora más poner la masa en movimiento. O bien, si el resorte es más duro, se va a mover más rápido y esto está bien: el periodo es menor si el resorte es más duro.

Noten que el periodo de oscilación de una masa fija a un resorte no depende en forma alguna de *cómo* se inició, de cuánto lo hemos estirado. El *periodo* está determinado, pero la amplitud de oscilación *no* está determinada por la ecuación de movimiento (21.2). La amplitud *queda* determinada, de hecho, por la manera en que lo soltamos, por lo que llamamos *condiciones iniciales* o condiciones de partida.

Realmente, no hemos encontrado completamente la solución más general posible de la ecuación (21.2). Hay otras soluciones. Debería estar claro el porqué: porque todos los casos representados por $x = a\cos\omega_0 t$ empiezan con un desplazamiento inicial y sin velocidad inicial. Pero es posible, por ejemplo, que la masa empiece en $x = 0$ y podríamos darle entonces un golpe impulsivo de manera que tenga alguna velocidad para $t = 0$. Este movimiento no está representado por un coseno —está representado por un seno—. Para expresarlo de otra manera, si $x = \cos\omega_0 t$ es una solución, entonces, ¿no es evidente que seguiría moviéndose de la misma manera que si entrásemos en la sala en un cierto tiempo (que llamaríamos "$t = 0$") y viéramos la masa en el momento que pasa por $x = 0$? Por lo tanto, $x = \cos\omega_0 t$ no puede ser la solución más general; debe ser posible correr el comienzo del tiempo, por decirlo así. Como un ejemplo, podríamos escribir la solución de esta manera: $x = a\cos\omega_0(t - t_1)$, donde t_1 es alguna constante. También esto corresponde a cambiar el origen del tiempo a un nuevo instante. Además, podemos desarrollar

$$\cos(\omega_0 t + \Delta) = \cos\omega_0 t\cos\Delta - \operatorname{sen}\omega_0 t \operatorname{sen}\Delta,$$

y escribir

$$x = A\cos\omega_0 t + B\operatorname{sen}\omega_0 t,$$

donde $A = a\cos\Delta$ y $B = -a\operatorname{sen}\Delta$. Cualquiera de estas formas es una manera posible de escribir la solución general completa de (21.2), es decir, toda solución de la ecuación diferencial $d^2x/dt^2 = -\omega_0^2 x$ que exista en el mundo se puede escribir

$$\text{(a)} \qquad x = a\cos\omega_0(t - t_1),$$

o

$$\text{(b)} \qquad x = a\cos(\omega_0 t + \Delta), \tag{21.6}$$

o

$$\text{(c)} \qquad x = A\cos\omega_0 t + B\,\mathrm{sen}\,\omega_0 t.$$

Algunas de las cantidades en (21.6) tienen nombres: ω_0 se llama *frecuencia angular*; es el número de radianes en el cambio de fase en un segundo. Eso queda determinado por la ecuación diferencial. Las otras constantes no están determinadas por la ecuación, sino por la manera en que comenzó el movimiento. De estas constantes, a mide el desplazamiento máximo alcanzado por la masa y se llama *amplitud* de la oscilación. La constante Δ se llama a veces *fase* de la oscilación, pero esto es una confusión porque otras personas llaman fase a $\omega_0 t + \Delta$, y dicen que la fase cambia con el tiempo. Podríamos decir que Δ es un *desfase* respecto a algún cero definido. Expresémoslo de otra manera. Las constantes Δ diferentes corresponden a movimientos con fases diferentes. Esto es cierto, pero si queremos o no llamar a Δ *la* fase, esto es otra cosa.

21-3 Movimiento armónico y movimiento circular

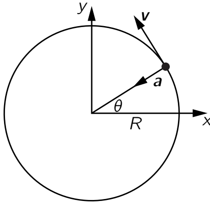


Figura 21-2. Una partícula moviéndose en una trayectoria circular con velocidad constante.

El hecho de que cosenos aparezcan en la solución de la ecuación (21.2) sugiere que a lo mejor hay alguna relación con círculos. Esto es artificial por supuesto, porque realmente no hay ningún círculo implicado en el movimiento rectilíneo: sencillamente va hacia arriba y hacia abajo. Podemos indicar que hemos, de hecho, resuelto ya esa ecuación diferencial, cuando estábamos estudiando la mecánica del movimiento circular. Si una partícula se mueve en un círculo a velocidad constante v, el vector radio desde el centro del círculo a la partícula gira en un ángulo cuyo tamaño es proporcional al tiempo. Si llamamos este ángulo $\theta = vt/R$ (figura 21-2) entonces $d\theta/dt = \omega_0 = v/R$. Sabemos que hay una aceleración $a = v^2/R = \omega_0^2 R$ hacia el centro. También sabemos que la posición x en un momento dado es el radio del círculo por $\cos\theta$ y que y es el radio por $\mathrm{sen}\,\theta$:

$$x = R\cos\theta, \qquad y = R\,\mathrm{sen}\,\theta.$$

¿Qué pasa con la aceleración? ¿Qué es la componente x de la aceleración, d^2x/dr^2? Esto ya lo hemos resuelto geométricamente; es la magnitud de la aceleración por el coseno del ángulo de proyección, con un signo menos porque es hacia el centro.

$$a_x = -a\cos\theta = -\omega_0^2 R\cos\theta = -\omega_0^2 x. \tag{21.7}$$

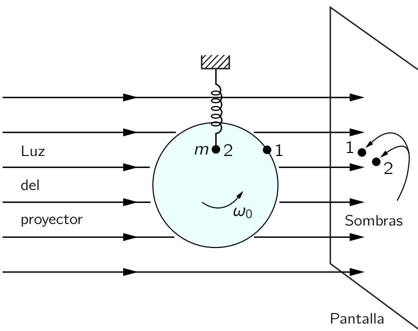


Figura 21-3. Demostración de la equivalencia entre movimiento armónico simple y movimiento circular uniforme.

En otras palabras, cuando una partícula se está moviendo en un círculo, la componente horizontal de su movimiento tiene una aceleración que es proporcional al desplazamiento horizontal desde el centro. Por supuesto, también tenemos la solución para el movimiento en un círculo: $x = R\cos\omega_0 t$. La ecuación (21.7) no depende del radio del círculo; de manera que para un círculo de cualquier radio se encuentra la misma ecuación para un ω_0 dado. Luego, por varias razones, esperamos que el desplazamiento de una masa fija a un resorte resultará ser proporcional a $\cos\omega_0 t$ y va a ser, en realidad, exactamente el mismo movimiento que el que veríamos si observáramos la componente x de la posición de un objeto que estuviera rotando en un círculo con velocidad angular ω_0. Como una verificación de esto, uno puede diseñar un experimento para demostrar que el movimiento hacia arriba y hacia abajo de una masa fija a un resorte es el mismo que el de un punto que da vueltas en un círculo. En la figura 21-3 una luz de arco proyectada sobre una pantalla forma las sombras de un perno de un volante solidario a un eje y de una masa que oscila verticalmente, uno al lado de otro. Si soltamos la masa en el tiempo preciso desde la posición precisa, y si la velocidad

del eje se ajusta cuidadosamente de manera que las frecuencias coincidan, cada uno debería seguir al otro exactamente. También se puede comparar la solución numérica que obtuvimos antes con la función coseno y ver si concuerdan muy bien.

Aquí podemos indicar que, dado que el movimiento uniforme en un círculo está tan relacionado matemáticamente con el movimiento oscilatorio hacia arriba y hacia abajo, podemos analizar el movimiento oscilatorio de una manera más simple, si lo imaginamos como la proyección de algo que se mueve en un círculo. En otras palabras, aunque la distancia y no significa nada en el problema oscilatorio, de todos modos podemos complementar la ecuación (21.2) con otra ecuación en y y juntarlas. Si hacemos esto, podremos analizar nuestro oscilador en una dimensión con movimientos *circulares*, que es mucho más fácil que tener que resolver una ecuación diferencial. El truco para hacer esto es usar números complejos, un procedimiento que vamos a introducir en el próximo capítulo.

21-4 Condiciones iniciales

Ahora veamos qué es lo que determina las constantes A y B o a y Δ. Éstas están determinadas, por supuesto, por la manera en que empezamos el movimiento. Si empezamos el movimiento simplemente con un pequeño desplazamiento, esto da un tipo de oscilación; si empezamos con un desplazamiento inicial y empujamos hacia arriba al soltar, obtenemos otro movimiento. Las constantes A y B o a y Δ, o cualquier otra manera de expresarlas, están determinadas, por supuesto, por la manera en que comenzó el movimiento y no por ningún otro aspecto de la situación. Éstas se llaman *condiciones iniciales*. Nos gustaría relacionar las condiciones iniciales con las constantes. Aunque esto puede hacerse usando cualquiera de las formas (21.6), resulta más sencillo si usamos la ecuación (21.6 c). Supongan que para $t = 0$ hemos empezado con un desplazamiento inicial x_0 y una cierta velocidad v_0. Ésta es la manera más general en que podemos empezar el movimiento. (Es cierto que no podemos especificar la *aceleración* con que empezó, porque eso queda determinado por el resorte, una vez que especificamos x_0.) Ahora, calculemos A y B. Empezamos con la ecuación para x,

$$x = A\cos\omega_0 t + B\,\mathrm{sen}\,\omega_0 t.$$

Como más adelante vamos a necesitar la velocidad también, derivamos x y obtenemos

$$v = -\omega_0 A\,\mathrm{sen}\,\omega_0 t + \omega_0 B\cos\omega_0 t.$$

Estas expresiones son válidas para todo t, pero tenemos un conocimiento específico respecto a x y v para $t = 0$. De manera que si ponemos $t = 0$ en estas ecuaciones, obtenemos x_0 y v_0 porque ése es el valor de x y de v para $t = 0$; también sabemos que el coseno de cero es uno y el seno de cero es cero. Por lo tanto, obtenemos

$$x_0 = A\cdot 1 + B\cdot 0 = A$$

y

$$v_0 = -\omega_0 A\cdot 0 + \omega_0 B\cdot 1 = \omega_0 B.$$

De manera que para este caso particular encontramos

$$A = x_0, \qquad B = v_0/\omega_0.$$

De estos valores de A y B podemos obtener a y Δ si queremos.

Esto es el final de nuestra solución, pero existe un hecho físico interesante a verificar y éste es la conservación de la energía. Como no hay pérdidas por fricción, la energía debería conservarse. Usemos la fórmula

$$x = a\cos(\omega_0 t + \Delta);$$

luego

$$v = -\omega_0 a \operatorname{sen}(\omega_0 t + \Delta).$$

Averigüemos ahora cuál es la energía cinética T y cuál es la energía potencial U. La energía potencial en cualquier momento es $\frac{1}{2}kx^2$, donde x es el desplazamiento y k es la constante del resorte. Si sustituimos x, usando la expresión anterior, obtenemos

$$U = \tfrac{1}{2}kx^2 = \tfrac{1}{2}ka^2 \cos^2(\omega_0 t + \Delta).$$

La energía potencial no es constante por supuesto; el potencial nunca se hace negativo naturalmente —siempre hay alguna energía en el resorte, pero la cantidad de energía fluctúa con x—. La energía cinética, por otra parte, es $\frac{1}{2}mv^2$ y sustituyendo v obtenemos

$$T = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m\omega_0^2 a^2 \operatorname{sen}^2(\omega_0 t + \Delta).$$

Ahora bien, la energía cinética es cero cuando x es máxima, porque entonces ahí no hay velocidad; por otro lado, es máxima cuando x pasa por cero, porque entonces se está moviendo más rápido. La variación de la energía cinética es justamente opuesta a la de la energía potencial. Pero la energía total debería ser constante. Si observamos que $k = m\omega_0^2$, vemos que

$$T + U = \tfrac{1}{2}m\omega_0^2 a^2[\cos^2(\omega_0 t + \Delta) + \operatorname{sen}^2(\omega_0 t + \Delta)] = \tfrac{1}{2}m\omega_0^2 a^2.$$

La energía depende del cuadrado de la amplitud; si tenemos una amplitud doble, obtenemos una oscilación con una energía cuatro veces mayor. La energía potencial *media* es la mitad del máximo y, por lo tanto, la mitad del total y la energía cinética media también es la mitad de la energía total.

21-5 Oscilaciones forzadas

A continuación vamos a discutir el *oscilador armónico forzado*, es decir, uno en el cual actúa una fuerza motriz externa. La ecuación es entonces la siguiente:

$$m\, d^2x/dt^2 = -kx + F(t). \tag{21.8}$$

Nos gustaría averiguar qué sucede en estos casos. La fuerza motriz externa puede tener diversos tipos de dependencia funcional en el tiempo; la primera que vamos a analizar es muy simple —vamos a suponer que la fuerza está oscilando—:

$$F(t) = F_0 \cos \omega t. \tag{21.9}$$

Noten, sin embargo, que este ω no es necesariamente ω_0: tenemos ω bajo nuestro control. Se puede forzar a diferentes frecuencias. Así que tratamos de resolver la ecuación (21.8) con la fuerza particular (21.9). ¿Cuál es la solución de (21.8)? Una solución particular (vamos a discutir luego los casos más generales) es

$$x = C \cos \omega t, \tag{21.10}$$

donde la constante debe ser determinada. En otras palabras, podemos suponer que si seguimos empujando hacia atrás y hacia adelante, la masa va a seguir hacia atrás y hacia adelante al compás de la fuerza. De todos modos podemos tratarlo. Así que introducimos (21.10) y (21.9) en (21.8) y obtenemos:

$$-m\omega^2 C \cos \omega t = -m\omega_0^2 C \cos \omega t + F_0 \cos \omega t. \tag{21.11}$$

También hicimos $k = m\omega_0^2$, para que entendamos mejor la ecuación al final. Ahora bien, como los cosenos aparecen en todas partes, los podemos simplificar y eso muestra que (21.10) es en realidad una solución con tal que elijamos C correctamente. La respuesta es que C debe ser

$$C = F_0/m(\omega_0^2 - \omega^2). \tag{21.12}$$

Esto es, m oscila a la misma frecuencia que la fuerza, pero con una amplitud que depende de la frecuencia de la fuerza y también de la frecuencia del movimiento natural del oscilador. Significa, primero, que si ω es muy pequeño comparado con ω_0, entonces el desplazamiento y la fuerza están en la misma dirección. Por otro lado, si lo sacudimos muy rápido de una parte a otra, entonces (21.12) nos dice que C es negativa si ω está por sobre la frecuencia natural ω_0 del oscilador armónico. (Llamaremos a ω_0 la frecuencia natural del oscilador armónico y ω la frecuencia aplicada.) A muy alta frecuencia el denominador puede hacerse muy grande, entonces no hay mucha amplitud.

Por supuesto que la solución que hemos encontrado es una solución sólo si las cosas comienzan correctamente, porque de otra manera hay una parte que por lo común desaparece después de un tiempo. Esta otra parte se llama *respuestas transitorias* a $F(t)$, mientras que (21.10) y (21.12) se llaman respuesta de *régimen estacionario*.

De acuerdo con nuestra fórmula (21.12), una cosa notable debe ocurrir también: si ω es casi igual a ω_0, entonces C debe tender a infinito. De manera que si ajustamos la frecuencia de la fuerza para que esté "a tiempo" con la frecuencia natural, entonces deberíamos obtener un enorme desplazamiento. Esto es bien conocido para cualquier persona que haya empujado un niño en un columpio. No da resultado que cerremos los ojos y que empujemos con una cierta velocidad al azar. Si lo hacemos en el momento oportuno, el columpio sube muy alto; pero si estamos a un ritmo malo, a veces podríamos estar empujando cuando deberíamos estar tirando, etc., y la cosa no resulta.

Si hacemos ω exactamente igual a ω_0 encontramos que debería oscilar con una amplitud *infinita*, lo que, por supuesto, es imposible. La razón por la cual no lo hace es que en la ecuación hay algo malo, hay otros términos de fricción y otras fuerzas que no aparecen en (21.8), pero que existen en el mundo real. De manera que la amplitud no llega a infinito por alguna razón; ¡podría ser que el resorte se rompiera!

22

Álgebra

22-1 Adición y multiplicación

En nuestro estudio de los sistemas oscilatorios vamos a tener la oportunidad de usar una de las fórmulas más notables, casi asombrosa, de toda la matemática. Desde el punto de vista del físico podríamos introducir esta fórmula en dos minutos más o menos y estar listos. Pero la ciencia existe tanto para el goce intelectual como para la utilidad práctica, de manera que en vez de demorarnos unos cuantos minutos en esta maravillosa joya, vamos a colocar la joya dentro de su marco apropiado en el grandioso diseño de aquella rama de la matemática que se llama álgebra elemental.

Ahora ustedes podrían preguntar: "¿Qué hace la matemática en una clase de física?" Tenemos varias excusas posibles; primero, por supuesto, la matemática es una herramienta importante, pero ésta solamente nos excusaría por entregar la fórmula en dos minutos. Por otra parte, en física teórica descubrimos que todas nuestras leyes pueden ser escritas en forma matemática, y que esto tiene cierta sencillez y belleza. Así, en último término para entender la naturaleza podría ser necesario tener una comprensión más profunda de las relaciones matemáticas. Pero la razón verdadera es que el tema es entretenido y aunque nosotros los humanos dividimos la naturaleza de maneras diferentes y tenemos cursos diferentes en departamentos distintos, esta departamentalización es realmente artificial y deberíamos disfrutar nuestros placeres intelectuales donde los encontremos.

Otra razón para examinar con mayor cuidado el álgebra, aunque la mayoría de nosotros estudió álgebra en el colegio, es que ésa fue la primera vez que la estudiábamos; todas las ecuaciones eran poco familiares y era muy difícil, así como lo es la física ahora. De vez en cuando es un gran placer mirar hacia atrás para ver qué territorio ha sido estudiado y cuál es el gran mapa o plan de todo el conjunto. ¡Quizás, algún día, alguien en el Departamento de Matemáticas va a dar una clase de mecánica de tal manera que muestre qué era lo que estábamos tratando de aprender en el curso de física!

El tema de álgebra no va a ser desarrollado desde el punto de vista del matemático, exactamente, porque los matemáticos están interesados más que nada en cómo se muestran los diversos hechos matemáticos, en cuántas suposiciones son absolutamente necesarias y en lo que no es necesario. No están tan interesados en el resultado de lo que demuestran. Por ejemplo, podemos encontrar el teorema de Pitágoras muy interesante, que la suma de los cuadrados de los lados de un triángulo rectángulo es igual al cuadrado de la hipotenusa; ése es un hecho interesante, una cosa curiosamente simple, que puede ser apreciada sin discutir el hecho de cómo probarlos o qué axiomas se necesitan. Así, en el mismo espíritu vamos a describir cualitativamente, si es que podemos hacerlo así, el sistema del álgebra elemental. Decimos álgebra *elemental* porque hay una rama de la matemática llamada álgebra *moderna* en la cual algunas de las reglas como $ab = ba$ se abandonan y todavía se sigue llamando álgebra, pero no vamos a discutir esto.

Para discutir este tema, empezamos por el medio. Suponemos que ya sabemos lo que son los enteros, lo que es el cero y lo que significa aumentar un número en una unidad. Ustedes podrían decir: "¡Eso no está en el medio!" Pero eso es el medio desde un punto de vista matemático porque podríamos ir más atrás aún y describir la teoría de los conjuntos para *deducir* alguna de estas propiedades de los enteros. Pero no vamos en esa dirección, la dirección de la filosofía matemática

y de la lógica matemática, sino más bien en la otra dirección a partir de la suposición de que sabemos lo que son los enteros y que sabemos contar.

Si empezamos con un cierto número a, un entero, y contamos sucesivamente una unidad b veces, el número a que llegamos lo llamamos $a+b$, y eso define la *adición* de enteros.

Una vez que hemos definido la adición, podemos considerar esto: si empezamos con nada y le sumamos a, b veces sucesivamente, llamamos el resultado *multiplicación* de enteros: lo llamamos b *por* a.

También podemos tener una *sucesión de multiplicaciones*: si empezamos con 1 y multiplicamos por a, b veces sucesivamente, llamamos a esto *elevar a una potencia*: a^b.

Ahora, como consecuencia de estas definiciones, se puede demostrar fácilmente que todas las siguientes relaciones son verdaderas:

$$\begin{array}{llll} \text{(a)} & a+b=b+a & \text{(b)} & a+(b+c)=(a+b)+c \\ \text{(c)} & ab=ba & \text{(d)} & a(b+c)=ab+ac \\ \text{(e)} & (ab)c=a(bc) & \text{(f)} & (ab)^c=a^cb^c \\ \text{(g)} & a^ba^c=a^{(b+c)} & \text{(h)} & (a^b)^c=a^{(bc)} \\ \text{(i)} & a+0=a & \text{(j)} & a\cdot 1=a \\ \text{(k)} & a^1=a & & \end{array} \tag{22.1}$$

Estos resultados son muy conocidos y no vamos a insistir en ellos, sólo los indicamos. Por supuesto que 1 y 0 tienen propiedades especiales; por ejemplo, $a+0=a$, $a\times 1=a$, y a elevada a la primera potencia es a.

En esta discusión también debemos suponer algunas otras propiedades como continuidad y orden que son muy difíciles de definir; dejaremos que la teoría rigurosa lo haga. Además, es totalmente cierto que hemos escrito demasiadas "reglas"; algunas de ellas pueden ser deducidas de otras, pero no nos vamos a preocupar por estas cosas.

22-2 Las operaciones inversas

Además de las operaciones directas de adición, multiplicación y elevación a una potencia, también tenemos las operaciones *inversas* que se definen como sigue: supongamos que a y c están dados y que queremos encontrar cuáles valores de b satisfacen ecuaciones como $a+b=c$, $ab=c$, $b^a=c$. Si $a+b=c$, b se define como $c-a$ que se llama *sustracción.* La operación llamada división también es clara: si $ab=c$, entonces $b=c/a$ define la división —una solución de la ecuación $ab=c$ "hacia atrás"—. Si tenemos una potencia $b^a=c$ y nos preguntamos "¿Qué es b?", b se llama la *raíz* a-ésima de c: $b=\sqrt[a]{c}$. Por ejemplo, si nos hacemos la siguiente pregunta: "¿qué entero, elevado a la tercera potencia, es igual a 8?", entonces la respuesta se llama *raíz cúbica* de 8; es 2. Como b^a y a^b no son iguales, hay *dos* problemas inversos asociados con las potencias y el otro problema inverso debe ser: "¿A qué potencia debemos elevar 2 para obtener 8?" Esto se llama tomar el *logaritmo.* Si $a^b=c$, escribimos $b=\log_a c$. El hecho que tenga una notación más complicada en relación con los otros no significa que sea menos elemental, por lo menos aplicado a los enteros, que los otros procesos. Aunque los logaritmos aparecen tarde en un curso de álgebra, por cierto que en la práctica son tan sencillos como las raíces; son precisamente una solución diferente de una ecuación algebraica. Las operaciones directas e inversas se resumen como sigue:

$$\begin{array}{llll} \text{(a)} & \text{adición} & \text{(a}'\text{)} & \text{sustracción} \\ & a+b=c & & b=c-a \\ \text{(b)} & \text{multiplicación} & \text{(b}'\text{)} & \text{división} \\ & ab=c & & b=c/a \\ \text{(c)} & \text{potencia} & \text{(c}'\text{)} & \text{raíz} \\ & b^a=c & & b=\sqrt[a]{c} \\ \text{(d)} & \text{potencia} & \text{(d}'\text{)} & \text{logaritmo} \\ & a^b=c & & b=\log_a c \end{array} \tag{22.2}$$

Ahora, he aquí la idea. Estas relaciones o reglas son correctas para enteros, ya que se deducen de las definiciones de adición, multiplicación y elevación a potencia. *Vamos a discutir si podemos o no ampliar la clase de objetos que representan a, b y c, de manera que obedezcan estas mismas reglas*, aunque los procesos para $a + b$, etc., no van a ser definibles en términos de la acción directa de sumar l, por ejemplo, o multiplicación sucesiva por enteros.

22-3 Abstracción y generalización

Cuando tratamos de resolver ecuaciones algebraicas simples, usando todas estas definiciones, pronto descubrimos algunos problemas insolubles como el siguiente. Supongan que tratamos de resolver la ecuación $b = 3 - 5$. Esto significa, de acuerdo con nuestra definición de sustracción, que debemos encontrar un número que cuando se suma a 5 da 3. Y por supuesto no existe tal número, porque consideramos sólo enteros positivos; éste es un problema insoluble. Sin embargo, el plan, la gran idea es ésta: *abstracción y generalización*. De la estructura completa del álgebra, reglas más enteros, abstraemos las definiciones originales de adición y multiplicación, pero dejamos las reglas (22.1) y (22.2) y suponemos que éstas son válidas *en general* para una clase de números más amplia, aunque ellas fueron obtenidas originalmente en una clase menor. Así, en vez de usar enteros simbólicamente para definir las reglas, usamos las reglas como definición de los símbolos, los que entonces representan un tipo de número más general. Como ejemplo, trabajando solamente con las reglas, podemos demostrar que $3 - 5 = 0 - 2$. De hecho podemos demostrar que uno puede hacer *todas* las sustracciones siempre que definamos un conjunto completo de nuevos números: $0 - 1, 0 - 2, 0 - 3, 0 - 4$, etc., llamados *enteros negativos*. Entonces podemos usar todas las otras reglas, como $a(b + c) = ab + ac$, etc., para encontrar cuáles son las reglas para multiplicar números negativos, y vamos a descubrir que todas las reglas pueden ser mantenidas tanto con enteros negativos como positivos.

De manera que hemos aumentado el rango de los objetos para los cuales las reglas dan resultado, pero el significado de los símbolos es diferente.

Uno no puede decir, por ejemplo, que -2 por 5 realmente significa 5 sucesivamente -2 veces. Eso no tiene ningún significado. Sin embargo, todo va a resultar bien de acuerdo con las reglas.

Un problema interesante aparece al considerar las potencias. Supongan que queremos descubrir lo que significa $a^{(3-5)}$. Sólo sabemos que $3-5$ es una solución del problema $(3 - 5) + 5 = 3$. Sabiendo eso, sabemos que $a^{(3-5)}a^5 = a^3$. Por lo tanto $a^{(3-5)} = a^3/a^5$ por la definición de división. Con un poco más de trabajo, esto puede ser reducido a $1/a^2$. De manera que encontramos que las potencias negativas son los recíprocos de las potencias positivas, pero $1/a^2$ es un símbolo sin significado porque si a es un entero positivo o negativo, su cuadrado puede ser mayor que 1 y ¡todavía no sabemos lo que entendemos por 1 dividido entre un número mayor que 1!

¡Adelante! El gran plan es continuar el proceso de generalización; cada vez que encontramos otro problema que no podemos resolver extendemos nuestro reino de los números. Consideren la división: no podemos encontrar un número que sea un entero, aunque sea negativo, que sea igual al resultado de dividir 3 entre 5. Pero si suponemos que todos los números fraccionarios también satisfacen las reglas, entonces podemos hablar de multiplicar y sumar fracciones y todo resulta tan bien como antes.

Consideren otro ejemplo de potencias: ¿qué es $a^{3/5}$? Sólo sabemos que $(3/5)5 = 3$, ya que esa fue la definición de $3/5$. Razón por la cual sabemos que $(a^{(3/5)})^5 = a^{(3/5)(5)} = a^3$ porque ésta es una de las reglas. Entonces por la definición de raíces encontramos que $a^{(3/5)} = \sqrt[5]{a^3}$.

De esta manera podemos definir lo que queremos decir al poner fracciones en lugar de los diversos símbolos, usando las reglas mismas para ayudarnos a determinar la definición —no es arbitrario—. ¡Es un hecho sorprendente que todas las reglas sigan válidas tanto para los enteros positivos y negativos como para las fracciones!

Sigamos con el proceso de generalización. ¿Hay alguna otra ecuación que no podamos resolver? Sí, hay otra. Por ejemplo, es imposible resolver esta ecuación: $b = 2^{1/2} = \sqrt{2}$. Es imposible encontrar un número que sea racional (una fracción) cuyo cuadrado sea igual a 2. Es muy fácil para nosotros en los tiempos modernos responder a esta pregunta. Conocemos el sistema decimal, de manera que no tenemos dificultad en entender el significado de un decimal sin término como un tipo de aproximación a la raíz cuadrada de 2. Históricamente, esta idea presentó una gran dificultad para los griegos. Para definir *en forma precisa* lo que se quiere decir aquí, es necesario agregar alguna esencia de continuidad y orden, y esto es, de hecho, exactamente en este punto casi el paso más difícil en el proceso de generalización. Fue hecho formal y rigurosamente por Dedekind. Sin embargo, sin preocuparnos del rigor matemático del asunto es muy fácil de entender que lo que queremos decir es que vamos a encontrar una sucesión completa de fracciones aproximadas, fracciones perfectas (porque cualquier decimal, cuando se corta en alguna parte, es ciertamente racional), que siguen y siguen, acercándose cada vez más al resultado deseado. Esto es suficiente para lo que queremos discutir y ello permite enredarnos con los números irracionales y calcular con bastante esfuerzo cosas como la raíz cuadrada de 2 con toda la precisión que deseamos.

22-4 Cómo obtener valores aproximados de números irracionales

El siguiente problema surge cuando averiguamos qué sucede con las potencias irracionales. Supongan que queremos definir, por ejemplo, $10^{\sqrt{2}}$. En principio la respuesta es bastante sencilla. Si aproximamos la raíz cuadrada de 2 a un cierto número de cifras decimales, entonces la potencia es racional y podemos extraer la raíz aproximada, usando el método anterior, y obtener una *aproximación* para $10^{\sqrt{2}}$. Después le podemos agregar unas cuantas cifras decimales más (de nuevo ella es racional), extraer la raíz apropiada, esta vez una raíz de orden mucho mayor porque hay un denominador mayor en la fracción, y obtener una mejor aproximación. Por supuesto, vamos a tener unas raíces enormes y el trabajo es bastante difícil. ¿Cómo vamos a superar este problema?

En el cálculo de las raíces cuadradas, raíces cúbicas y otras raíces pequeñas hay un proceso aritmético mediante el cual podemos obtener una cifra decimal después de la otra. Pero la cantidad de trabajo necesario para calcular potencias irracionales y los logaritmos que van con ellas (el problema inverso) es tan grande que no hay ningún proceso aritmético simple que podamos usar. Por lo tanto, se han construido tablas que nos permiten calcular estas potencias; y éstas se llaman tablas de logaritmos o tablas de potencias dependiendo de la manera que están construidas. Es sencillamente un problema de ahorrar tiempo; si debemos elevar un número a una potencia irracional es mejor buscarlo que tener que calcularlo. Por supuesto, este cálculo no es más que un problema técnico, pero interesante y de gran valor histórico. En primer lugar, no sólo tenemos el problema de resolver $x = 10^{\sqrt{2}}$, sino también tenemos el problema de resolver $10^x = 2$, o $x = \log_{10} 2$. Éste no es un problema en el cual tengamos que definir un nuevo tipo de número para el resultado, es simplemente un problema de cálculo. La respuesta es simplemente un número irracional, un término decimal, no un nuevo tipo de número.

Discutamos ahora el problema de calcular soluciones de tales ecuaciones. La idea general es verdaderamente muy simple. Si pudiéramos calcular 10^1 y $10^{4/10}$ y $10^{1/100}$y $10^{4/1000}$, etc., y los multiplicáramos, obtendríamos $10^{1.414\ldots}$, o $10^{\sqrt{2}}$, y ésta es la idea general de cómo esto funciona. Pero en vez de calcular $10^{1/10}$ y los demás, vamos a calcular $10^{1/2}$, $10^{1/4}$, etc. Antes que empecemos, debemos explicar por qué hacemos tanto trabajo con 10, en vez de con otro número. Nos damos cuenta, por supuesto, que las tablas de logaritmos son de gran utilidad práctica, aparte del problema matemático de extraer raíces, ya que con cualquier base

$$\log_b(ac) = \log_b a + \log_b c. \tag{22.3}$$

Todos sabemos que podemos usar este hecho de una manera práctica para multiplicar números si tenemos una tabla de logaritmos. La única pregunta es:

¿con qué base b vamos a calcular? No importa qué base se use; podemos usar el mismo principio todas las veces que sea y si estamos usando logaritmos con una base determinada, podemos encontrar los logaritmos respecto a cualquier otra base simplemente cambiando la escala, un factor de multiplicación. Si multiplicamos la ecuación (22.3) por 61, sigue siendo igualmente válida; y si tuviéramos una tabla de logaritmos con base b y si alguien multiplicara toda nuestra tabla por 61, no habría diferencia esencial. Supongan que conocemos los logaritmos de todos los números en la base b. En otras palabras, podemos resolver la ecuación $b^a = c$ para cualquier c porque tenemos una tabla. El problema es encontrar el logaritmo del mismo número c con respecto a otra base, digamos x. Nos gustaría resolver $x^{a'} = c$. Ello es fácil de hacer, porque siempre podemos escribir $x = b^t$, que define a t, conociendo x y b. De hecho $t = \log_b x$. Si remplazamos este valor y despejamos para a', vemos que $(b^t)^{a'} = b^{a't} = c$. En otras palabras, ta' es el logaritmo de c en base b. Así $a' = a/t$. Luego los logaritmos en base x son simplemente $1/t$, que es una constante, por los logaritmos en base b. Por lo tanto, cualquier tabla de logaritmos es equivalente a cualquier otra tabla de logaritmos si multiplicamos por una constante y la constante es $1/\log_b x$. Esto nos permite elegir una base particular y por conveniencia tomamos la base 10. (La pregunta puede surgir de si existe alguna base natural, una base en la cual las cosas sean más sencillas y vamos a tratar de encontrar una respuesta a eso más tarde. Por el momento vamos a usar la base 10.)

Tabla 22-1

Raíces cuadradas sucesivas de 10

Exponente s	$1\,024\,s$	10^s	$(10^s - 1)/s$
1	1024	10.00000	9.00
1/2	512	3.16228	4.32
1/4	256	1.77828	3.113
1/8	128	1.33352	2.668
1/16	64	1.15478	2.476
1/32	32	1.074607	2.3874
1/64	16	1.036633	2.3445
1/128	8	1.018152	2.3234 ^{211}
1/256	4	1.0090350	2.3130 ^{104}
1/512	2	1.0045073	2.3077 ^{53}
1/1 024	1	1.0022511	2.3051 ^{26}
			↓ ^{26}
$\Delta/1\,024$ $(\Delta \to 0)$	Δ	$1 + 0.0022486\Delta$ ←	2.3025

Veamos ahora cómo calcular logaritmos. Empezamos calculando raíces cuadradas sucesivas de 10 por el método de aproximación. Los resultados se muestran en la tabla 22-1. La primera columna da las potencias de 10 y el resultado, 10^s, se da en la tercera columna. Así, $10^1 = 10$. La potencia a un medio de 10 la podemos calcular fácilmente porque es la raíz cuadrada de 10 y hay un proceso conocido y simple para extraer la raíz cuadrada de cualquier número.* Usando este proceso, encontramos que la primera raíz cuadrada es 3.16228. ¿De qué nos sirve eso? Ya nos dice algo, nos dice cómo encontrar $10^{0.5}$, así conocemos por lo menos *un* logaritmo, si sucede que necesitamos el logaritmo de 3.16228 sabemos que la respuesta está cerca de 0.50000. Pero tenemos que mejorar un poco; se observa que necesitamos más información. De manera que extraemos la raíz cuadrada de nuevo y encontramos $10^{1/4}$ que es 1.77828. Ahora tenemos el logaritmo de más números que antes, 1.250 es el logaritmo de 17.78 y, entre paréntesis, si alguien

* Existe un procedimiento aritmético definido, pero la manera más fácil de encontrar la raíz cuadrada de cualquier número N es elegir un a bastante cerca del valor, encontrar N/a, promediar $a' = \frac{1}{2}[a + (N/a)]$ y usar este valor promediado a' para la siguiente elección de a. La convergencia es muy rápida: el número de cifras significativas se dobla cada vez.

pregunta por $10^{0.75}$ lo podemos obtener porque es $10^{(0.5+0.25)}$; es, por lo tanto, el producto de los números segundo y tercero. Si podemos obtener números en la columna *s* para poder construir casi cualquier número, entonces al multiplicar los elementos adecuados en la columna 3, podemos obtener 10 elevado a cualquier potencia; ése es el plan. Así que calculamos 10 raíces cuadradas sucesivas y ése es el trabajo principal de estos cálculos.

¿Por qué no seguimos para obtener cada vez más precisión? Porque empezamos a darnos cuenta de algo. Cuando elevamos 10 a una potencia muy pequeña, obtenemos 1 más una pequeña cantidad. La razón de esto es clara porque vamos a tener que tomar la milésima potencia de $10^{1/1\,000}$ para volver a 10, de manera que mejor que no empecemos con un número tan grande; tiene que ser cerca de 1. Lo que notamos es que los pequeños números que sumamos a 1 empiezan a aparecer como si estuviéramos simplemente dividiendo entre 2 cada vez; vemos que 1 815 se hace 903; después 450, 225; de manera que está claro que si extraemos otra raíz, con excelente aproximación obtendremos 1.00112 más o menos, y en vez de *extraer* realmente todas las raíces cuadradas, *estimamos* el límite último. Cuando tomamos una pequeña fracción $\Delta/1\,024$, ¿cuál va a ser la respuesta a medida que Δ tienda a cero? Por supuesto, va a ser algún número cerca de $1 + 0.0022511\Delta$. No exactamente $1 + 0.0022511\Delta$; sin embargo, podemos obtener un valor mejor mediante el siguiente truco: restamos el 1 y después dividimos por la potencia *s*. Esto debería corregir todos los excesos en el mismo valor. Vemos que son muy parecidos. Al principio de la tabla no, pero a medida que bajan, se acercan cada vez más a un valor constante. ¿Cuál es el valor? De nuevo miramos cómo varía la serie, cómo ha cambiado con *s*. Cambió en 211, en 104, en 53, en 26. Estos cambios son evidentemente en forma muy aproximada la mitad el uno del otro a medida que bajamos. Por lo tanto, si seguimos adelante, los cambios serían 13, 7, 3, 2, 1, más o menos, o un total de 26. De manera que tenemos que avanzar solamente en 26 y así encontramos que el número verdadero es 2.3025 (realmente veremos más tarde que el número *exacto* debería ser 2.3026, pero para no quitarle realidad no vamos a alterar nada en la aritmética). A partir de esta tabla podemos calcular ahora cualquier potencia de 10, componiendo las potencias a partir de la 1024ª.

Calculemos ahora realmente un logaritmo, porque el proceso que vamos a usar es el proceso de donde provienen las tablas de logaritmos. El procedimiento se muestra en la tabla 22-2 y los valores numéricos se muestran en la tabla 22-1 (columnas 2 y 3).

Tabla 22-2

Cálculo de un logaritmo: $\log_{10} 2$

$$
\begin{aligned}
&2 \div 1.77828 = 1.124682\\
&1.124682 \div 1.074607 = 1.046598, \text{ etc.}\\
\therefore\, &2 = (1.77828)(1.074607)(1.036633)(1.0090350)(1.000573)\\
&= 10^{\left[\frac{(256+32+16+4+0.254)}{1\,024}\right]} = 10^{\left[\frac{308.254}{1\,024}\right]}\\
&= 10^{0.30103} \qquad \left(\frac{573}{2\,249} = 0.254\right)\\
\therefore\, &\log_{10} 2 = 0.30103
\end{aligned}
$$

Supongan que queremos el logaritmo de 2. Es decir, queremos saber a qué potencia debemos elevar 10 para obtener 2. ¿Podemos elevar 10 a la potencia 1/2? No; es muy grande. En otras palabras, podemos ver que la respuesta va a ser mayor que 1/4 y menor que 1/2. Saquemos el factor $10^{1/4}$; dividimos 2 entre 1.778 . . . , y obtenemos 1.124 . . . , etc., y ahora sabemos que hemos sacado 0.250000 del logaritmo. El número 1.124. . . , es ahora el número cuyo logaritmo necesitamos. Cuando hayamos terminado, agregaremos de nuevo el 1/4 o 256/1 024. Ahora buscamos en la tabla el número siguiente, justo debajo de 1.124 . . . , y es 1.074607.

Por lo tanto, dividimos entre 1.074607 y obtenemos 1.046598. De ahí descubrimos que 2 puede ser formado por un producto de números que están en la tabla 22-1, como sigue:

$$2 = (1.77828)(1.074607)(1.036633)(1.0090350)(1.000573).$$

Sobró un factor (1.000573), naturalmente, que está más allá del alcance de nuestra tabla. Para obtener el logaritmo de este factor usamos nuestro resultado que $10^{\Delta/1\,024} \approx 1 + 2.3025\Delta/1\,024$. Encontramos $\Delta = 0.254$. Luego nuestra respuesta es 10 a la siguiente potencia: $(256 + 32 + 16 + 4 + 0.254)/1\,024$. Sumándolos obtenemos $308.254/1\,024$. Dividiendo obtenemos 0.30103, de manera que sabemos que $\log_{10} 2 = 0.30103$; ¡que resulta correcto hasta 5 cifras!

Ésta es la manera como los logaritmos fueron calculados originalmente por el señor Briggs de Halifax, en 1620. Dijo, "computé sucesivamente 54 raíces cuadradas de 10". Sabemos que realmente calculó las primeras 27, porque el resto se puede obtener por este truco con Δ. Su trabajo consistió en calcular 27 veces la raíz cuadrada de 10, que no es mucho más que las 10 veces que lo hicimos nosotros; sin embargo, fue mucho más trabajo porque él calculó hasta 16 cifras decimales y luego redujo su resultado a 14 cuando lo publicó, de manera que no había errores de redondeo. Hizo tablas de logaritmos con 14 cifras decimales mediante este método, que es bastante tedioso. Pero por 300 años todas las tablas de logaritmos fueron tomadas de las tablas del señor Briggs, reduciendo el número de cifras decimales. Sólo en los tiempos contemporáneos, con la WPA y las máquinas computadoras, se han calculado nuevas tablas independientemente. Hay métodos mucho más eficientes de calcular logaritmos hoy en día, usando ciertos desarrollos en serie.

En el proceso anterior descubrimos algo bastante interesante, y es que para pequeñas potencias ϵ podemos calcular 10^{ϵ} fácilmente; hemos descubierto que $10^{\epsilon} = 1 + 2.3025\epsilon$ por simple análisis numérico. Esto también significa, por supuesto, que $10^{n/2.3025} = 1 + n$ si n es muy pequeña. Ahora bien, los logaritmos en cualquier otra base son simples múltiplos de los logaritmos en base 10. La base 10 fue usada solamente porque tenemos 10 dedos y la aritmética es fácil, pero si preguntamos por una base matemáticamente natural, una que no tenga nada que ver con el número de dedos en los seres humanos, podríamos tratar de cambiar nuestra *escala* de logaritmos de una manera natural y conveniente y el método que la gente ha elegido es redefinir los logaritmos multiplicando todos los logaritmos en base 10 por $2.3025\ldots$ Esto corresponde entonces a usar otra base y que se llama base *natural* o base e. Noten que $\log_e(1 + n) \approx n$ o $e^n \approx 1 + n$ a medida que $n \to 0$.

Es bastante fácil averiguar lo que e es, $e = 10^{1/2.3025\ldots}$ o $e = 10^{0.434310\ldots}$, una potencia irracional. Nuestra tabla de raíces cuadradas sucesivas de 10 se puede usar para calcular no sólo logaritmos, sino también 10 elevado a cualquier potencia; así que usémosla para calcular esta base natural e. Por conveniencia transformamos $0.434310\ldots$ en $444.73/1\,024$. Ahora, 444.73 es $256 + 128 + 32 + 16 + 8 + 4 + 0.73$. Entonces e, ya que es un exponente de una suma, va a ser el producto de los números

$$(1.77828)(1.33352)(1.074607)(1.036633)(1.018152)(1.009035)(1.001643) = 2.7184.$$

(El único problema es el último, que es 0.73, y que no está en la tabla, pero sabemos que si Δ es pequeño, la respuesta es $1 + 0.0022486\Delta$.) Cuando los multiplicamos, obtenemos 2.7184 (debería ser 2.7183, pero está bastante bien). El uso de estas tablas es, entonces, la manera mediante la cual se calculan las potencias irracionales y los logaritmos de números irracionales. Y con esto terminamos con los números irracionales.

22-5 Números complejos

Ahora resulta que, después de todo ese trabajo, ¡*todavía* no podemos resolver todas las ecuaciones! Por ejemplo, ¿cuál es la raíz cuadrada de -1? Supongan

que tenemos que encontrar $x^2 = -1$. El cuadrado de ningún racional, de ningún irracional, de *nada* que hayamos descubierto hasta ahora es igual a -1. De manera que de nuevo tenemos que generalizar nuestros números a una clase aún más amplia. Supongamos que una solución específica de $x^2 = -1$ se llama de alguna manera: nosotros la llamaremos i; i tiene por definición la propiedad de que su cuadrado es -1. Esto es calculando lo que vamos a decir de ella; por supuesto que hay más de una raíz de la ecuación $x^2 = -1$. Alguien podría escribir i, pero otro podría decir: "No, yo prefiero $-i$. Mi i es tu menos i". Es una solución igualmente buena, y como la única definición que tiene i es que $i^2 = -1$ debe ser cierto que cualquier ecuación que podamos escribir es igualmente cierta, si el signo de i se cambia por doquier. Esto se llama tomar el *complejo conjugado*. Ahora vamos a formar números sumando sucesivas i y multiplicando i por números y sumando otros números, etc., de acuerdo con todas nuestras reglas. De esta manera encontramos que los números pueden escribirse todos en la forma $p + iq$, donde p y q son los que llamamos números *reales*, es decir, los números que hemos estado definiendo hasta ahora. El número i se llama *unidad imaginaria*. Cualquier múltiplo real de i se llama *imaginario puro*. El número más general a es de la forma $p + iq$ y se llama *número complejo*. Las cosas no se empeoran si, por ejemplo, multiplicamos dos de estos números, digamos $(r + is)(p + iq)$. Entonces, usando las reglas, obtenemos:

$$\begin{aligned}(r + is)(p + iq) &= rp + r(iq) + (is)p + (is)(iq)\\ &= rp + i(rq) + i(sp) + (ii)(sq)\\ &= (rp - sq) + i(rq + sp),\end{aligned} \tag{22.4}$$

ya que, $ii = i^2 = -1$. Por lo tanto, todos los números que obedecen ahora las reglas (22.1) tienen esta forma matemática.

Ahora ustedes dirán: "¡Esto puede seguir indefinidamente! Hemos definido las potencias de los imaginarios y todo lo demás y cuando estamos listos, viene alguien con otra ecuación que no puede ser resuelta como $x^6 + 3x^2 = -2$. ¡Entonces tenemos que generalizar todo de nuevo!" Pero resulta que *con esta invención adicional* que es simplemente la raíz cuadrada de -1, *¡toda ecuación algebraica puede ser resuelta!* Éste es un hecho fantástico que debemos dejar que lo demuestre el Departamento de Matemáticas. Las demostraciones son hermosas y muy interesantes, pero ciertamente no son evidentes por sí mismas. De hecho, la suposición más evidente es que vamos a tener que inventar de nuevo, de nuevo y de nuevo. Pero el milagro más grande es que no tenemos que hacerlo. Ésta es la última invención. Después de esta invención de los números complejos, encontramos que las reglas siguen funcionando con los números complejos y hemos terminado de inventar cosas nuevas. Podemos encontrar la potencia compleja de cualquier número complejo, podemos resolver cualquier ecuación escrita algebraicamente en términos de un número finito de esos símbolos. No encontramos más números nuevos. La raíz cuadrada de i, por ejemplo, tiene un resultado bien determinado, no es algo nuevo; e i^i es algo. Vamos a discutir esto ahora.

Ya hemos discutido la multiplicación, y la adición también es sencilla; si sumamos dos números complejos $(p+iq)+(r+is)$ la respuesta es $(p+r)+i(q+s)$. Ya podemos sumar y multiplicar números complejos. Pero el problema real, por supuesto, es calcular potencias complejas de números complejos. Resulta que el problema es realmente no mucho más difícil que calcular las *potencias complejas de números reales*. Así que concretémonos ahora al problema de calcular 10 elevado a una potencia compleja, no precisamente una potencia irracional, sino $10^{(r+is)}$. Por supuesto, en todo momento debemos usar nuestras reglas (22.1) y (22.2). Así,

$$10^{(r+is)} = 10^r 10^{is}. \tag{22.5}$$

Pero ya sabemos cómo calcular 10^r y siempre podemos multiplicar algo por cualquier otra cosa; luego el problema es calcular solamente 10^{is}. Indiquémoslo por algún número complejo $x + iy$. Problema: dado s, encontrar x y encontrar y.

Ahora, si

$$10^{is} = x + iy,$$

entonces el complejo conjugado de esta ecuación también debe ser cierto, de manera que

$$10^{-is} = x - iy.$$

(Vemos así que podemos deducir una cantidad de cosas sin realmente calcular nada, utilizando nuestras reglas.) Deducimos otra cosa interesante al multiplicarlas:

$$10^{is}10^{-is} = 10^0 = 1 = (x+iy)(x-iy) = x^2 + y^2. \tag{22.6}$$

De manera que si encontramos x también encontramos y.

Ahora el problema es *cómo* calcular 10 elevado a una potencia imaginaria. ¿Qué guía existe? Podemos trabajar basándonos en nuestras reglas hasta que no podamos ir más allá, pero hay una guía razonable: si podemos calcularlo para un exponente s particular podemos obtenerlo para todo el resto. Si conocemos 10^{is} con una s cualquiera y después lo queremos para el doble de esa s, podemos elevar el número al cuadrado, y así sucesivamente. Pero ¿cómo encontrar 10^{is} aun para un valor especial de s? Para lograrlo vamos a hacer una hipótesis adicional que no está precisamente en la categoría de todas las otras reglas, pero que lleva a resultados razonables y nos permite progresar: cuando la potencia es pequeña vamos a suponer que la "ley" $10^\epsilon = 1 + 2.3025\epsilon$ es correcta, a medida que ϵ se hace muy pequeña, no sólo para ϵ real, sino *también para* ϵ *complejo.* Por lo tanto, empezamos con la hipótesis de que esta ley es válida en general, y esto nos dice que $10^{is} = 1 + 2.3025 \cdot is$, cuando $s \to 0$. De manera que suponemos que si s es muy pequeña, digamos uno en 1 024, tenemos una buena aproximación a 10^{is}.

Ahora hacemos una tabla mediante la cual podemos calcular *todas* las potencias imaginarias de 10, esto es calcular x y y. Esto se hace como sigue. La primera potencia con que empezamos es la potencia $1/1\,024$, que suponemos que tiene un valor muy cercano a $1 + 2.3025i/1\,024$. Así empezamos con:

$$10^{i/1\,024} = 1.00000 + 0.0022486i, \tag{22.7}$$

y si seguimos multiplicando el número por sí mismo, podemos llegar a una potencia imaginaria mayor. De hecho, basta con invertir el procedimiento que usamos al hacer nuestra tabla de logaritmos y calcular el cuadrado, la 4^a, la 8ª potencia, etc., de (22.7) y así construir los valores mostrados en la tabla 22-3. Notamos un hecho interesante, que los números x son positivos al principio, pero después se hacen negativos. Vamos a examinar esto un poco más dentro de un momento. Pero primero podríamos tener la curiosidad de averiguar para qué número s la parte real de 10^{is} es *cero.* El valor y podría ser 1 y así tendríamos $10^{is} = 1i$ o $is = \log_{10} i$. Como ejemplo de cómo usar esta tabla, así como calculamos antes $\log_{10} 2$, usamos la tabla 22-3 para encontrar $\log_{10} i$.

¿Cuál de los números de la tabla 22-3 debemos multiplicar para obtener un resultado imaginario? Después de tantear un poco, descubrimos que para reducir x lo más que se pueda, es mejor multiplicar "512" por "128". Esto da $0.13056 + 0.99159i$. Entonces descubrimos que debemos multiplicar esto por un número cuya parte imaginaria sea casi igual a la dimensión de la parte real que estamos tratando de eliminar. Por lo tanto, elegimos "64" cuyo valor de y es 0.14349, ya que es el más cercano a 0.13056. Esto da $-0.01308 + 1.00008i$. Ahora nos pasamos de largo y debemos *dividir* entre $0.99996 + 0.00900i$. ¿Cómo lo hacemos? Cambiando el signo de i y multiplicando por $0.99996 - 0.00900i$ (que funciona si $x^2 + y^2 = 1$). Siguiendo de esta manera encontramos que la potencia a que debemos elevar 10 para que dé i es $i(512 + 128 + 64 - 4 - 2 + 0.20)/1\,024$, o $698.20i/1\,024$. Si elevamos 10 a esa potencia obtenemos i. Luego $\log_{10} i = 0.68184i$.

Tabla 22-3

Cuadrados sucesivos de $10^{i/1\,024} = 1 + 0.0022486i$

Exp is	$1\,024s$	10^{is}
$i/1\,024$	1	$1.00000 + 0.00225i$*
$i/512$	2	$1.00000 + 0.00450i$
$i/256$	4	$0.99996 + 0.00900i$
$i/128$	8	$0.99984 + 0.01800i$
$i/64$	16	$0.99936 + 0.03599i$
$i/32$	32	$0.99742 + 0.07193i$
$i/16$	64	$0.98967 + 0.14349i$
$i/8$	128	$0.95885 + 0.28402i$
$i/4$	256	$0.83872 + 0.54467i$
$i/2$	512	$0.40679 + 0.91365i$
$i/1$	1 024	$-0.66928 + 0.74332i$
	* Debe ser $0.0022486i$	

22-6 Exponentes imaginarios

Tabla 22-4

Potencias sucesivas de $10^{i/8}$

$p =$ potencia $\cdot\, 8/i$	$10^{ip/8}$
0	$1.00000 + 0.00000i$
1	$0.95882 + 0.28402i$
2	$0.83867 + 0.54465i$
3	$0.64944 + 0.76042i$
4	$0.40672 + 0.91356i$
5	$0.13050 + 0.99146i$
6	$-0.15647 + 0.98770i$
7	$-0.43055 + 0.90260i$
8	$-0.66917 + 0.74315i$
9	$-0.85268 + 0.52249i$
10	$-0.96596 + 0.25880i$
11	$-0.99969 - 0.02620i$
12	$-0.95104 - 0.30905i$
14	$-0.62928 - 0.77717i$
16	$-0.10447 - 0.99453i$
18	$+0.45454 - 0.89098i$
20	$+0.86648 - 0.49967i$
22	$+0.99884 + 0.05287i$
24	$+0.80890 + 0.58836i$

Para seguir investigando el tema de elevar a potencias complejas imaginarias, echemos un vistazo a las potencias de 10 al elevar a *potencias sucesivas*, no duplicando cada vez la potencia con el fin de continuar con la tabla 22-3 y ver qué pasa con esos signos menos. Esto se muestra en la tabla 22-4 en la cual tomamos $10^{i/8}$ y lo seguimos multiplicando. Vemos que x disminuye, pasa por cero, llega casi a -1 (si pudiéramos introducirnos entre $p = 10$ y $p = 11$ evidentemente alcanzaría -1) y vuelve atrás. El valor y también va de un lado a otro.

En la figura 22-1, los puntos representan los números que aparecen en la tabla 22-4 y las líneas se dibujaron precisamente para ayudarles visualmente. Vemos así que los números x y y oscilan; 10^{is} *se repite*, es algo *periódico* y como tal es fácil de explicar, porque si una cierta potencia es i, entonces la cuarta potencia de ello será i^2 al *cuadrado*. Será $+1$ de nuevo y por lo tanto, ya que $10^{0.68i}$ es igual a i, al tomar la cuarta potencia descubrimos que $10^{2.72i}$ es igual a $+1$. Por lo tanto, si quisiéramos $10^{3.00i}$, por ejemplo, podríamos escribirlo como $10^{2.72i}$ por $10^{0.28i}$. En otras palabras, tiene un periodo, se repite. ¡Por supuesto, reconocemos a qué se parecen las curvas! Se parecen al seno y al coseno y las vamos a llamar, mientras tanto, el seno algebraico y el coseno algebraico. Sin embargo, en vez de usar la base 10, las vamos a poner en nuestra base natural, lo que cambia solamente la escala horizontal; de manera que remplazamos $2.3025s$ por t y escribimos $10^{is} = e^{it}$, donde t es un número real, Ahora $e^{it} = x + iy$, y lo vamos a escribir como el coseno algebraico de t más i veces el seno algebraico de t. Así,

$$e^{it} = \underline{\cos}\, t + i\, \underline{\operatorname{sen}}\, t. \tag{22.8}$$

¿Cuáles son las propiedades de $\underline{\cos}\, t$ y $\underline{\operatorname{sen}}\, t$? Sabemos primero, por ejemplo, que $x^2 + y^2 = 1$; lo hemos demostrado antes y es tan válido para la base e como para la base 10. Luego $\underline{\cos}^2\, t + \underline{\cos}^2\, t = 1$. También sabemos que para t pequeño, $e^{it} = 1 + it$ y, por lo tanto $\underline{\cos}\, t$ es casi 1 y $\underline{\operatorname{sen}}\, t$ es casi t, y así resulta que *todas las diversas propiedades de estas notables funciones*, que provienen de elevar a potencias imaginarias, *son las mismas que las de seno y coseno de la trigonometría.*

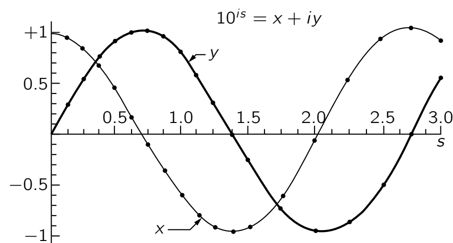


Figura 22-1

¿Es el periodo el mismo? Averigüémoslo. ¿Elevado a qué potencia e es igual a i? ¿Cuál es el logaritmo de i en base e? Ya resolvimos esto anteriormente; en la base 10 era $0.68184i$, pero cuando cambiamos nuestra escala logarítmica a e, tenemos que multiplicar por 2.3025, y si lo hacemos resulta 1.570. De manera que lo vamos a llamar "$\pi/2$ algebraico". Pero vemos que se diferencia del $\pi/2$ regular sólo por una unidad en la última cifra; ¡y eso, por supuesto, es el resultado de errores en nuestra aritmética! Por lo tanto, hemos creado dos funciones nuevas con una manera puramente algebraica, el coseno y el seno, que pertenecen al álgebra y sólo al álgebra. Despertamos al final descubriendo las mismas funciones que son propias de la geometría. De manera que aquí hay una conexión, en última instancia, entre el álgebra y la geometría.

Resumimos con esto la fórmula más notable de la matemática:

$$e^{i\theta} = \cos\theta + i \operatorname{sen}\theta. \tag{22.9}$$

Ésta es nuestra joya.

Podemos relacionar la geometría con el álgebra, representando los números complejos en un plano; la posición horizontal de un punto es x, la posición vertical de un punto es y (figura 22-2). Representar cualquier número complejo $x + iy$. Luego si la distancia radial a este punto la llamamos r y el ángulo se designa por θ, la ley algebraica es que $x + iy$ se escribe de la forma $re^{i\theta}$, donde las relaciones geométricas entre x, y, r y θ son como se indican. Ésta es, entonces, la unificación del álgebra y la geometría.

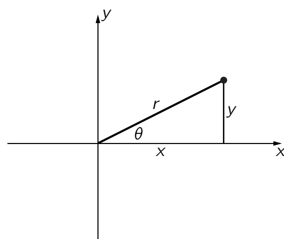


Figura 22-2. $x + iy = re^{i\theta}$.

Cuando empezamos este capítulo, provistos sólo con las nociones básicas de enteros y de contar, no teníamos idea del poder del proceso de abstracción y generalización. Usando el conjunto de "leyes" algebraicas o propiedades de los números, ecuación (22.1), y las definiciones de operaciones inversas (22.2), hemos

sido capaces, nosotros mismos, de fabricar no sólo números, sino cosas útiles como tablas de logaritmos, potencias y funciones trigonométricas (porque esto es lo que son las potencias imaginarias de números reales), todo ¡simplemente extrayendo 10 raíces cuadradas sucesivas de 10!

23

Resonancia

23-1 Números complejos y movimiento armónico

En este capítulo vamos a continuar nuestra discusión del oscilador armónico y en particular el oscilador armónico forzado, usando una nueva técnica en el análisis. En el capítulo anterior introdujimos la idea de números complejos, que tienen partes real e imaginaria y que pueden ser representados en un diagrama en el cual la ordenada representa la parte imaginaria y la abscisa representa la parte real. Si a es un número complejo, podemos escribirlo como $a = a_r + ia_i$ donde el subíndice r significa la parte real de a y el subíndice i significa la parte imaginaria de a. Refiriéndonos a la figura 23-1 vemos que también podemos escribir un número complejo $a = x + iy$; en la forma $x + iy = re^{i\theta}$, donde $r^2 = x^2 + y^2 = (x + iy)(x - iy) = aa^*$. (El complejo conjugado de a, escrito a^*, se obtiene cambiando el signo de i en a.) Vamos así a representar un número complejo en cualquiera de las dos formas, una parte real más una parte imaginaria o un módulo r y un así llamado ángulo de fase θ. Dado r y θ, x y y son claramente $r\cos\theta$ y $r\operatorname{sen}\theta$ e inversamente, dado un número complejo $x + iy, r = \sqrt{x^2 + y^2}$ y $\tan\theta = y/x$, el cociente entre la parte imaginaria y la real.

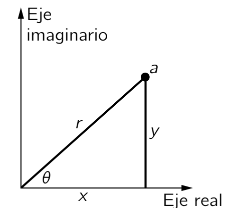


Figura 23-1. Un número complejo se puede representar con un punto en el “plano complejo”.

Vamos a aplicar números complejos a nuestro análisis de fenómenos físicos mediante el siguiente truco. Tenemos ejemplos de cosas que oscilan; la oscilación puede tener una fuerza impulsora, que es una cierta constante por $\cos\omega t$. Ahora, esta fuerza $F = F_0\cos\omega t$ se puede escribir como la parte real de un número complejo $F = F_0 e^{i\omega t}$, porque $e^{i\omega t} = \cos\omega t + i\operatorname{sen}\omega t$. La razón por la cual hacemos esto es que es más fácil trabajar con una función exponencial que con coseno. De manera que todo el truco es representar nuestras funciones oscilantes como partes reales de ciertas funciones complejas. El número complejo F que hemos definido así no es una fuerza física real, porque ninguna fuerza en física es realmente compleja, las fuerzas reales no tienen partes imaginarias, sólo una parte real. Vamos, sin embargo, a hablar de la “fuerza” $F_0 e^{i\omega t}$, pero por supuesto la fuerza verdadera es la *parte real* de esa expresión.

Tomemos otro ejemplo. Supongan que queremos representar una fuerza que es una onda cosinusoidal, que está desfasada con un atraso de fase Δ. Ésta sería, por supuesto, la parte real de $F_0 e^{i(\omega t - \Delta)}$, pero siendo las exponenciales lo que son, podemos escribir $e^{i(\omega t - \Delta)} = e^{i\omega t}e^{-i\Delta}$. Vemos así que el álgebra de las exponenciales es mucho más fácil que la de los senos y cosenos, ésta es la razón por la que elegimos el uso de números complejos. Vamos a escribir muy a menudo

$$F = F_0 e^{-i\Delta} e^{i\omega t} = \hat{F} e^{i\omega t}. \tag{23.1}$$

Escribimos un acento circunflejo (ˆ) sobre F para recordarnos que esta cantidad es un número complejo, aquí el número es:

$$\hat{F} = F_0 e^{-i\Delta}.$$

Ahora resolvamos una ecuación usando números complejos para ver si podemos resolver un problema en algún caso real. Por ejemplo, tratemos de resolver

$$\frac{d^2x}{dt^2} + \frac{kx}{m} = \frac{F}{m} = \frac{F_0}{m}\cos\omega t, \tag{23.2}$$

donde F es la fuerza que impulsa al oscilador y x el desplazamiento. Ahora, aunque parezca absurdo, supongamos que x y F son realmente números complejos, esto con un propósito matemático solamente. Es decir, x tiene una parte real

y una parte imaginaria por i, y F tiene una parte real y una parte imaginaria por i. Ahora, si tuviéramos una solución de (23.2) con números complejos y sustituyéramos los números complejos en la ecuación, obtendríamos:

$$\frac{d^2(x_r + ix_i)}{dt^2} + \frac{k(x_r + ix_i)}{m} = \frac{F_r + iF_i}{m}$$

o

$$\frac{d^2x_r}{dt^2} + \frac{kx_r}{m} + i\left(\frac{d^2x_i}{dt^2} + \frac{kx_i}{m}\right) = \frac{F_r}{m} + \frac{iF_i}{m}.$$

Ahora bien, ya que si dos números complejos son iguales sus partes reales deben ser iguales, *y* sus partes imaginarias deben ser iguales, *deducimos que la parte real de x satisface la ecuación con la parte real de la fuerza.* Debemos hacer énfasis, sin embargo, que esta separación en una parte real y una parte imaginaria *no* es válida en general, sino que es válida sólo para ecuaciones que son *lineales*, es decir, para ecuaciones en las cuales x aparece en cada término sólo en primera potencia o potencia cero. Por ejemplo, si hubiera en la ecuación un término λx^2, entonces, cuando sustituyéramos $x_r + ix_i$ obtendríamos $\lambda(x_r + ix_i)^2$, pero cuando separáramos en partes real e imaginaria esto daría $\lambda(x_r^2 - x_i^2)$ como la parte real y $2i\lambda x_r x_i$ como la parte imaginaria. Vemos así que la parte real de la ecuación no incluiría sólo λx_r^2 sino también $-\lambda x_i^2$. En este caso obtenemos una ecuación diferente que la queríamos resolver, mezclada con x_i, esa cosa completamente artificial que introdujimos en nuestro análisis.

Probemos ahora nuestro nuevo método para el problema del oscilador forzado, que ya sabemos resolver. Queremos resolver la ecuación (23.2) como antes, pero decimos que vamos a tratar de resolver

$$\frac{d^2x}{dt^2} + \frac{kx}{m} = \frac{\hat{F}e^{i\omega t}}{m}, \tag{23.3}$$

donde $\hat{F}e^{i\omega t}$ es un número complejo. También x va a ser compleja por supuesto, pero recuerden la regla, tomen la parte real para averiguar lo que está pasando realmente. De manera que tratamos de resolver (23.3) para la solución forzada, vamos a discutir después otras soluciones. La solución forzada tiene la misma frecuencia que la fuerza aplicada y tiene una cierta amplitud de oscilación y una cierta fase, de manera que también se puede representar con algún número complejo $\hat{x}$, cuyo módulo representa el balanceo de x y cuya fase representa el atraso en el tiempo de la misma manera que para la fuerza. Ahora bien, un aspecto maravilloso de una función exponencial es que $d(\hat{x}e^{i\omega t})/dt = i\omega\hat{x}e^{i\omega t}$. Cuando derivamos una función exponencial, bajamos el exponente como un simple factor. La segunda derivada hace lo mismo, baja otro $i\omega$ de manera que es muy sencillo escribir inmediatamente, por sola inspección, la ecuación para $\hat{x}$; cada vez que vemos una derivación simplemente multiplicamos por $i\omega$. (¡La derivación es ahora tan fácil como la multiplicación! Esta idea de usar exponenciales en ecuaciones diferenciales lineales es casi tan grande como la invención de los logaritmos, donde la multiplicación se remplaza por la suma. Aquí la derivación se remplaza por la multiplicación. Así, nuestra ecuación resulta:

$$(i\omega)^2\hat{x} + (k\hat{x}/m) = \hat{F}/m. \tag{23.4}$$

(Hemos simplificado el factor común $e^{i\omega t}$.) ¡Vean lo simple que es! Ecuaciones diferenciales se convierten inmediatamente a simple vista en meras ecuaciones algebraicas, prácticamente tenemos la solución a simple vista, a saber:

$$\hat{x} = \frac{\hat{F}/m}{(k/m) - \omega^2},$$

Ya que $(i\omega)^2 = -\omega^2$. Esto se puede simplificar un poco sustituyendo $k/m = \omega_0^2$, lo que da

$$\hat{x} = \hat{F}/m(\omega_0^2 - \omega^2). \tag{23.5}$$

Ésta, por supuesto, es la solución que teníamos antes, porque ya que $m(\omega_0^2 - \omega^2)$ es un número real, los ángulos de fase de $\hat{F}$ y de $\hat{x}$ son los mismos (o a lo mejor 180° si $\omega^2 > \omega_0^2$), como se había dicho previamente. El módulo de $\hat{x}$, que mide hasta dónde oscila, está relacionado con el tamaño de $\hat{F}$ mediante el factor $1/m(\omega_0^2 - \omega^2)$, y este factor se hace enorme cuando ω es casi igual a ω_0. Así obtenemos una respuesta muy fuerte cuando aplicamos la frecuencia apropiada (si sostenemos un péndulo al extremo de una cuerda y lo sacudimos con la frecuencia adecuada, lo podemos hacer balancearse muy alto).

23-2 El oscilador forzado amortiguado

Así es, entonces, como analizamos el movimiento oscilatorio con la técnica matemática más elegante. Pero la elegancia de la técnica de ninguna manera se manifiesta en un problema que se puede resolver fácilmente por otros métodos. Se manifiesta sólo cuando se aplica a problemas más difíciles. Resolvamos, por lo tanto, otro problema más difícil, que además agrega un aspecto un poco más real al anterior. La ecuación (23.5) nos dice que si la frecuencia ω fuera exactamente igual a ω_0, obtendríamos una respuesta infinita. En realidad esta respuesta infinita no se produce, por supuesto, porque otras cosas, como el roce, que hemos ignorado hasta ahora, limitan la respuesta. Agreguemos entonces a la ecuación (23.2) un término de roce.

Ordinariamente tal problema es muy difícil debido al carácter y a la complejidad del término de roce. Sin embargo, hay muchas circunstancias en las cuales la fuerza de roce es *proporcional a la velocidad* con que se mueve el objeto. Un ejemplo de este roce es el roce en el movimiento lento de un objeto en aceite o líquido espeso. No existe fuerza cuando está en reposo, pero mientras más rápido se mueve más rápido debe pasar el aceite junto al objeto y mayor es la resistencia. Por lo tanto, vamos a suponer que hay, además de los términos en (23.2), otro término, una fuerza de resistencia proporcional a la velocidad: $F_f = -c\,dx/dt$. Va a ser conveniente en nuestro análisis matemático, escribir la constante c como m por γ para simplificar un poco la ecuación. Éste es precisamente el mismo truco que usamos con k, cuando lo remplazamos por $m\omega_0^2$ para simplificar el álgebra. Luego, nuestra ecuación va a ser:

$$m(d^2x/dt^2) + c(dx/dt) + kx = F \tag{23.6}$$

o, escribiendo $c = m\gamma$ y $k = m\omega_0^2$ y dividiendo por la masa m,

$$(d^2x/dt^2) + \gamma(dx/dt) + \omega_0^2 x = F/m. \tag{23.6a}$$

Ahora tenemos la ecuación en la forma más conveniente para resolverla. Si γ es muy pequeño, esto indica muy poco roce, si γ es muy grande, hay una enorme cantidad de roce. ¿Cómo resolvemos esta ecuación diferencial lineal? Supongan que la fuerza impulsora sea igual a $F_0 \cos(\omega t + \Delta)$, podríamos introducirla en (23.6a) y tratar de resolverla, pero en vez de eso, vamos a resolverla mediante nuestro nuevo método. Por lo tanto, escribimos F como la parte real de $\hat{F}e^{i\omega t}$ y x como la parte real de $\hat{x}e^{i\omega t}$ y los sustituimos en la ecuación (23.6a). Ni siquiera es necesario realizar la sustitución, porque podemos ver por inspección que la ecuación resultaría:

$$[(i\omega)^2\hat{x} + \gamma(i\omega)\hat{x} + \omega_0^2\hat{x}]e^{i\omega t} = (\hat{F}/m)e^{i\omega t}. \tag{23.7}$$

[De hecho, si tratáramos de resolver la ecuación (23.6a) por nuestro viejo método pedestre, realmente apreciaríamos la magia del método "complejo".] Si dividimos entre $e^{i\omega t}$ en ambos miembros, podemos obtener las respuestas $\hat{x}$ para una fuerza dada $\hat{F}$, ella es:

$$\hat{x} = \hat{F}/m(\omega_0^2 - \omega^2 + i\gamma\omega). \tag{23.8}$$

De manera que $\hat{x}$ de nuevo queda dada por $\hat{F}$ por un cierto factor. No hay nombre técnico para este factor, ninguna letra en particular para él, pero lo

podemos llamar R para fines de discusión:

$$R = \frac{1}{m(\omega_0^2 - \omega^2 + i\gamma\omega)}$$

y

$$\hat{x} = \hat{F}R. \tag{23.9}$$

(Aunque las letras γ y ω_0 son de uso muy común, esta R no tiene nombre particular.) Este factor R se puede escribir ya sea como $p + iq$ o como un cierto módulo ρ por ei^θ. Si se escribe como un cierto módulo por $e^{i\theta}$, veamos lo que significa. Ahora $\hat{F} = F_0e^{i\Delta}$, y la fuerza real F es la parte real de $F_0e^{i\Delta}e^{i\omega t}$, es decir, $F_0\cos(\omega t + \Delta)$. A continuación la ecuación (23.9) nos dice que $\hat{x}$ es igual a $\hat{F}R$. Así que escribiendo $R = \rho e^{i\theta}$ como otra expresión para R, obtenemos

$$\hat{x} = R\hat{F} = \rho e^{i\theta}F_0e^{i\Delta} = \rho F_0 e^{i(\theta+\Delta)}.$$

Finalmente, volviendo aún más atrás, vemos que la x física, que es la parte real de la $\hat{x}e^{i\omega t}$ compleja, es igual a la parte real de $\rho F_0e^{i(\theta+\Delta)}e^{i\omega t}$. Pero ρ y F_0 son reales y la parte real de $e^{i(\theta+\Delta+\omega t)}$ es simplemente $\cos(\omega t + \Delta + \theta)$. Así,

$$x = \rho F_0 \cos(\omega t + \Delta + \theta). \tag{23.10}$$

Esto nos dice que la amplitud de la respuesta es el módulo de la fuerza F multiplicada por un cierto factor de amplificación ρ, esto nos da la "cantidad" de oscilación. También nos dice, sin embargo, que x no está oscilando en fase con la fuerza que tiene una fase Δ, sino que está desfasada en una cantidad adicional θ. Por lo tanto, ρ y θ representan el tamaño de la respuesta y el desfase de la respuesta.

Ahora busquemos lo que es ρ. Si tenemos un número complejo, el cuadrado de su módulo es igual al número por su complejo conjugado, así

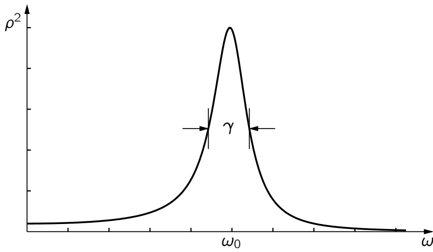


Figura 23-2. Gráfica de ρ^2 en función de ω.

$$\begin{aligned} \rho^2 &= \frac{1}{m^2(\omega_0^2 - \omega^2 + i\gamma\omega)(\omega_0^2 - \omega^2 - i\gamma\omega)} \\ &= \frac{1}{m^2[(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2]}. \end{aligned} \tag{23.11}$$

Además, el ángulo de fase θ es fácil de encontrar, porque si escribimos:

$$1/R = 1/\rho e^{i\theta} = (1/\rho)e^{-i\theta} = m(\omega_0^2 - \omega^2 + i\gamma\omega),$$

vemos que

$$\tan\theta = -\gamma\omega/(\omega_0^2 - \omega^2). \tag{23.12}$$

Es negativo $\tan(-\theta) = -\tan\theta$. Resulta un valor negativo para θ, para todo ω, y esto corresponde a que el desplazamiento x se retrasa respecto a la fuerza F.

La figura 23-2 muestra cómo ρ^2 varía en función de la frecuencia (ρ^2 es físicamente más interesante que ρ, porque ρ^2 es proporcional al cuadrado de la amplitud o, más o menos, a la *energía* que la fuerza desarrolla en el oscilador). Vemos que si γ es muy pequeño, entonces $1/(\omega_0^2 - w^2)^2$ es el término más importante y la respuesta tiende a ir a infinito cuando ω se hace igual a ω_0. Ahora, el "infinito" no es realmente infinito porque si $\omega = \omega_0$, todavía queda el $1/\gamma^2\omega^2$. El desfase varía como se muestra en la figura 23-3.

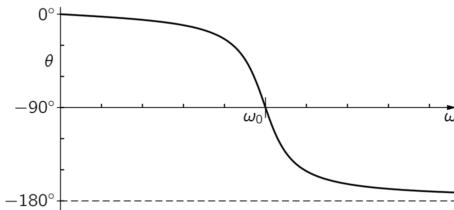


Figura 23-3. Gráfica de θ en función de ω.

En ciertas circunstancias, obtenemos una fórmula un poco diferente de (23.8), también llamada fórmula de resonancia y uno podría pensar que representa un fenómeno diferente, pero no es así. La razón es que si γ es muy pequeña, la parte más interesante de la curva está cerca de $\omega = w_0$ y podemos remplazar (23.8) por una fórmula aproximada que es muy precisa si γ es pequeña y ω está cerca de ω_0. Ya que $\omega_0^2 - \omega^2 = (\omega_0 - \omega)(\omega_0 + \omega)$, si ω está cerca de ω_0 esto es casi lo mismo que $2\omega_0(\omega_0 - \omega)$ y $\gamma\omega$ es casi lo mismo que $\gamma\omega_0$. Usando esto en (23.8), vemos que $\omega_0^2 - \omega^2 + i\gamma\omega \approx 2\omega_0(\omega_0 - \omega + i\gamma/2)$, de manera que

$$\hat{x} \approx \hat{F}/2m\omega_0(\omega_0 - \omega + i\gamma/2) \quad \text{si} \quad \gamma \ll \omega_0 \quad \text{y} \quad \omega \approx \omega_0. \tag{23.13}$$

Es fácil encontrar la fórmula correspondiente para ρ^2. Es

$$\rho^2 \approx 1/4m^2\omega_0^2[(\omega_0 - \omega)^2 + \gamma^2/4].$$

Vamos a dejar al estudiante que demuestre lo siguiente: si designamos por la unidad a la altura máxima de la curva de ρ^2 en función de ω y preguntamos por el ancho $\Delta\omega$ de la curva a la mitad de la altura máxima, el ancho total a mitad de la altura máxima de la curva es $\Delta\omega = \gamma$, suponiendo que γ es pequeño. La resonancia se hace cada vez más aguda a medida que los efectos de roce se hacen cada vez más pequeños.

Como otra medida del ancho, algunas personas usan una cantidad Q que se define como $Q = \omega_0/\gamma$. Mientras más angosta la resonancia, más alta la $Q : Q = 1\,000$ significa una resonancia de la que el ancho es sólo la milésima parte de la escala de frecuencia. La Q de la curva de resonancia mostrada en la figura 23-2 es 5.

La importancia del fenómeno de resonancia es que aparece en muchas otras circunstancias y, por lo tanto, el resto de este capítulo va a describir algunas de estas circunstancias.

23-3 Resonancia eléctrica

Las aplicaciones técnicas más simples y difundidas de la resonancia están en la electricidad. En el mundo eléctrico hay un cierto número de objetos que se pueden conectar para hacer circuitos eléctricos. Estos *elementos pasivos de circuito*, como se les llama comúnmente, son de tres tipos principales, aunque cada uno tiene un poco de los otros dos. Antes de describirlos en mayor detalle, notemos que nuestra idea acerca del oscilador mecánico se basa en una masa en el extremo de un resorte, pero esto es sólo una aproximación. Toda la masa no está realmente en la “masa”, algo de ella está en la inercia del resorte. Igualmente, el resorte no está en el “resorte”, la masa misma tiene un poco de elasticidad y aunque pudiera parecer, no es *absolutamente* rígida, y a medida que sube y baja, se deforma un poco bajo la acción del resorte que la está tirando. Lo mismo es válido en la electricidad. Hay una aproximación en la cual podemos agrupar cosas en “elementos de circuito” de los cuales se supone que tienen características puras, ideales. Éste no es el momento apropiado para discutir aquí esa aproximación; vamos a suponer sencillamente que es válida en este caso.

Los tres tipos principales de elementos de circuito son los siguientes. El primero se llama *capacitor* (figura 23-4), un ejemplo es dos placas planas metálicas separadas una muy pequeña distancia por un material aislador. Cuando las placas se cargan se produce entre ellas un voltaje, es decir, una cierta diferencia de potencial. La misma diferencia de potencial aparece entre los terminales A y B, porque si hubiera alguna diferencia entre los alambres de conexión la electricidad se escaparía. Por lo tanto, existe un cierto voltaje V entre las placas, si hay una cierta carga $+q$ y $-q$ en ellas, respectivamente. Entre las placas hay un cierto campo eléctrico, hasta hemos encontrado una fórmula para él (capítulos 13 y 14):

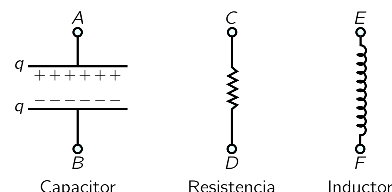


Figura 23-4. Los tres elementos pasivos de circuito.

$$V = \sigma d/\epsilon_0 = qd/\epsilon_0 A, \tag{23.14}$$

donde d es la separación y A el área de las placas. Noten que la diferencia de potencial es una función lineal de la carga. Si no tenemos placas paralelas, sino electrodos aislados de cualquier otra forma, la diferencia de potencial es todavía precisamente proporcional a la carga, pero la constante de proporcionalidad podría no ser tan fácil de calcular. Sin embargo, todo lo que necesitamos saber es que la diferencia de potencial a través de un condensador *es proporcional a la carga* $V = q/C$, la constante de proporcionalidad es $1/C$, donde C es la *capacitancia* del objeto.

El segundo tipo de elemento de circuito se llama *resistencia*, ofrece resistencia al paso de la corriente. Sucede que los alambres metálicos y muchas otras

sustancias resisten el paso de la corriente de esta manera, si hay una diferencia de potencial aplicada a un pedazo de alguna sustancia, existe una corriente eléctrica $I = dq/dt$ que es proporcional a la diferencia del potencial eléctrico:

$$V = RI = R\,dq/dt \tag{23.15}$$

El coeficiente de proporcionalidad se llama *resistencia* R. Esta relación puede ya ser familiar para ustedes, es la ley de Ohm.

Si pensamos que la carga q en un capacitor es análoga al desplazamiento x de un sistema mecánico, vemos que la corriente $I = dq/dt$ es análoga a la velocidad, $1/C$ es análoga a la constante k del resorte y R es análoga al coeficiente resistivo $c = m\gamma$ en la ecuación (23.6). Es muy interesante que ¡exista además otro elemento de circuito que es análogo a la *masa*! Es una bobina que forma un campo magnético dentro de sí misma cuando hay una corriente en ella. Un campo magnético *variable* desarrolla en la bobina un voltaje que es proporcional de dI/dt (de hecho, así funciona un transformador). El campo magnético es proporcional a la corriente y el voltaje inducido (así nombrado) en dicha bobina es proporcional a la rapidez de variación de la corriente:

$$V = L\,dI/dt = L\,d^2q/dt^2. \tag{23.16}$$

El coeficiente L es la *autoinductancia* y es análoga a la masa en un circuito mecánico oscilante.

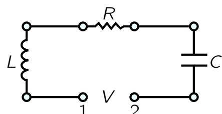


Figura 23-5. Un circuito eléctrico oscilatorio con resistencia, inductancia y capacitancia.

Supongan que hacemos un circuito en el cual hemos conectado tres elementos de circuito en serie (figura 23-5), luego el voltaje en los extremos del objeto de 1 a 2 es el trabajo realizado al llevar una carga a través de él y consiste en la suma de varias partes: en el inductor, $V_L = Ld^2q/dt^2$, en la resistencia $V_R = Rdq/dt$, en el capacitor $V_C = q/C$. Su suma es igual al voltaje aplicado V:

$$L\,d^2q/dt^2 + R\,dq/dt + q/C = V(t). \tag{23.17}$$

Ahora vemos que esta ecuación es exactamente igual a la ecuación mecánica (23.6), y por supuesto se puede resolver exactamente de la misma manera. Supongamos que $V(t)$ es oscilatorio, estamos impulsando el circuito con un generador y con una oscilación sinusoidal pura. Entonces podemos escribir $V(t)$ como un $\hat{V}$ complejo en el entendimiento que se debe multiplicar por $e^{i\omega t}$ en última instancia y luego se debe tomar la parte real para encontrar el verdadero V. Asimismo, se puede analizar la carga q y entonces, exactamente como en la ecuación (23.8), escribimos la ecuación correspondiente: la segunda derivada de $\hat{q}$ es $(i\omega)^2\hat{q}$; la primera derivada es $(i\omega)\hat{q}$. Por lo tanto, la ecuación (23.17) se traduce en:

$$\left[L(i\omega)^2 + R(i\omega) + \frac{1}{C}\right]\hat{q} = \hat{V}$$

o

$$\hat{q} = \frac{\hat{V}}{L(i\omega)^2 + R(i\omega) + \dfrac{1}{C}}$$

que podemos escribir en la forma

$$\hat{q} = \hat{V}/L(\omega_0^2 - \omega^2 + i\gamma\omega), \tag{23.18}$$

donde $\omega_0^2 = 1/LC$ y $\gamma = R/L$. ¡Es exactamente el mismo denominador que teníamos en el caso mecánico, con exactamente las mismas propiedades de resonancia! La correspondencia entre los casos eléctricos y mecánicos está delineada en la tabla 23-1.

Debemos mencionar un pequeño detalle técnico. En la literatura eléctrica se usa una notación diferente. (De un campo a otro el tema en realidad no es muy diferente, pero la manera de escribir las notaciones es a menudo diferente.) Primero, se usa comúnmente j en vez de i en ingeniería eléctrica para designar $\sqrt{-1}$. (¡Después de todo, i debe ser la corriente!) También los ingenieros prefieren

Tabla 23-1

Característica general	Propiedad mecánica	Propiedad eléctrica
variable indep.	tiempo (t)	tiempo (t)
variable dep.	posición (x)	carga (q)
inercia	masa (m)	inductancia (L)
resistencia	coef. de arrastre ($c = \gamma m$)	resistencia ($R = \gamma L$)
rigidez	rigidez (k)	$(\text{capacitancia})^{-1}$ $(1/C)$
frecuencia de reson.	$\omega_0^2 = k/m$	$\omega_0^2 = 1/LC$
periodo	$t_0 = 2\pi\sqrt{m/k}$	$t_0 = 2\pi\sqrt{LC}$
factor de mérito	$Q = \omega_0/\gamma$	$Q = \omega_0 L/R$

tener una relación entre $\hat{V}$ e $\hat{I}$ en vez de entre V y $\hat{q}$, simplemente porque están más acostumbrados a esa manera. Así, ya que $\hat{I} = d\hat{q}/dt = i\omega\hat{q}$, basta con sustituir $\hat{q}$ por $\hat{I}/i\omega$ y obtener

$$\hat{V} = (i\omega L + R + 1/i\omega C)\hat{I} = \hat{Z}\hat{I}. \tag{23.19}$$

Otra manera es reescribir la ecuación (23.17) de manera que parezca más familiar. Uno la ve a menudo escrita de esta manera:

$$L\,dI/dt + RI + (1/C)\int^t I\,dt = V(t). \tag{23.20}$$

En todo caso encontramos la relación (23.19) entre el voltaje $\hat{V}$ y la corriente $\hat{I}$, que es precisamente la misma que (23.18) excepto que está dividida entre $i\omega$ y esto produce la ecuación (23.19). La cantidad $R + i\omega L + 1/i\omega C$ es un número complejo y se usa tanto en ingeniería eléctrica que tiene un nombre, se llama *impedancia Complejas* $\hat{Z}$. Así, pues, podemos escribir $\hat{V} = \hat{Z}\hat{I}$. La razón por la cual a los ingenieros les gusta hacer esto es que aprendieron algo cuando eran jóvenes: $V = RI$ para las resistencias cuando sólo sabían de resistencia y CC. Ahora han llegado a ser más educados y tienen circuitos CA de manera que quieren que la ecuación se vea igual. Por eso escriben $\hat{V} = \hat{Z}\hat{I}$ con la única diferencia que la resistencia se remplaza por algo más complicado, una cantidad compleja. Así, pues, insisten en que no pueden usar lo que todo el resto del mundo usa para los números imaginarios, tienen que usar j para eso, ¡es un milagro que no insistieran también que la letra Z fuera una R! (Entonces se meten en líos cuando hablan de densidad de corriente, para la cual también usan j. Las dificultades de la ciencia son, en gran parte, las dificultades de las notaciones, las unidades y todas las otras artificialidades inventadas por el hombre y no por la naturaleza.)

23-4 Resonancia en la naturaleza

Aunque hemos discutido el caso eléctrico en detalle, podríamos presentar un caso tras otro en muchos campos y mostrar exactamente que la ecuación de resonancia es la misma. Existen muchas circunstancias en la naturaleza en donde algo está "oscilando" y sucede el fenómeno de resonancia. Lo dijimos en un capítulo anterior, demostrémoslo ahora. Si paseamos por nuestro estudio sacando libros de los estantes y simplemente los hojeamos para encontrar un ejemplo de una curva que corresponda a la figura 23-2 y que provenga de la misma ecuación, ¿qué encontramos? Precisamente para demostrar el amplio campo abarcado al tomar la muestra más pequeña posible se necesitan tomar sólo cinco o seis libros para obtener toda una serie de fenómenos que muestran resonancia.

Los dos primeros son de la mecánica, el primero en gran escala, la atmósfera de toda la Tierra. Si la atmósfera, que suponemos que rodea la Tierra en forma pareja por todos lados, es atraída hacia un lado por la Luna, o más bien aplastada alargándola en una doble marea y si pudiéramos después soltarla se pondría a chapotear de arriba a abajo, es un oscilador. Este oscilador es *impulsado* por la Luna, la que está, efectivamente, dando vueltas alrededor de la Tierra, cualquier

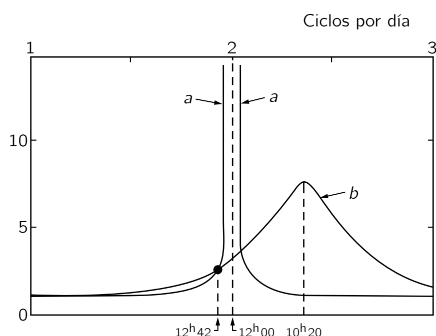


Figura 23-6. Respuesta de la atmósfera a una excitación externa. El valor de *a* es la respuesta adecuada si la marea atmosférica S_2 es de origen gravitacional; la amplificación del máximo es 100 : 1. El valor de *b* proviene de la amplificación y fase observadas para la marea M_2. [Munk y MacDonald, "Rotation of the Earth", Cambridge University Press (1960).]

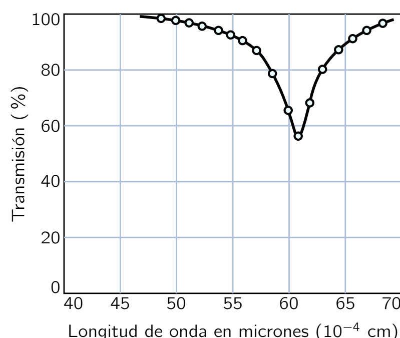


Figura 23-7. Transmisión de radiación infrarroja a través de una delgada (0.17 μ) película de cloruro de sodio. [De R. B. Barnes, Z. *Physik* **75**, 723 (1932). Kittel, *Introducción a la física del estado sólido*, Reverté, Barcelona, 1965.

componente de la fuerza, digamos en la dirección x, tiene una componente coseno, de manera que la respuesta de la atmósfera terrestre a la atracción de marea de la Luna es la de un oscilador. La respuesta esperada de la atmósfera se muestra en la figura 23-6 curva b (la curva a es otra curva teórica discutida en el libro de donde hemos sacado esto). Podría pensarse que tenemos sólo un punto en esta curva de resonancia, ya que tenemos sólo aquella frecuencia que corresponde a la rotación de la Tierra bajo la Luna que ocurre en un periodo de 12.42 horas —12 horas para la Tierra (la marea es un doble chichón) y un poco más porque la Luna está dando vueltas—. Pero, a partir del *tamaño* de las mareas atmosféricas y de la *fase*, el monto del retraso, podemos obtener tanto ρ como θ. De éstos podemos obtener ω_0 y γ y así ¡dibujar la curva entera! Éste es un ejemplo de ciencia muy pobre. De dos números obtenemos dos números y a partir de estos dos números dibujamos una hermosa curva que, por supuesto, ¡pasa por el mismo punto que determinó la curva! Esto no vale nada *a menos que podamos medir otra cosa* y en el caso de la geofísica eso es a menudo muy difícil. Pero en este caso particular existe otra cosa de la cual podemos demostrar teóricamente que debe tener el mismo sincronismo que la frecuencia natural ω_0, o sea, si alguien perturbara la atmósfera, ésta oscilaría con una frecuencia ω_0. Ahora bien, *hubo* una tal perturbación intensa en 1883, el volcán Krakatoa hizo explosión y la mitad de la isla salió volando, e hizo una explosión tan tremenda en la atmósfera que se pudo medir el periodo de oscilación de la atmósfera. Resultó ser $10\frac{1}{2}$ de horas. El ω_0 que se obtuvo de la figura 23-6 resulta 10 horas y 20 minutos, de manera que por lo menos tenemos una prueba de la realidad de nuestra comprensión de las mareas atmosféricas.

A continuación pasamos a las oscilaciones mecánicas en pequeña escala. Esta vez tomamos un cristal de cloruro de sodio, que tiene iones sodio y iones cloro uno junto al otro como lo describimos en un capítulo anterior. Estos iones están cargados eléctricamente, alternando positivos con negativos. Ahora es posible una oscilación interesante. Supongan que pudiéramos mover todas las cargas positivas hacia la derecha y todas las cargas negativas hacia la izquierda y soltarlas, entonces oscilarían de un lado para el otro, la red de sodio contra la red de cloro. ¿Cómo podemos alguna vez inducir tal cosa? Esto es fácil, porque si aplicamos un campo eléctrico al cristal ¡va a empujar las cargas positivas hacia un lado y las negativas hacia el otro! Así, teniendo un campo eléctrico externo podremos a lo mejor obtener que el cristal oscile. Sin embargo, ¡la frecuencia del campo eléctrico necesaria es tan alta, que corresponde a la *radiación infrarroja*! De manera que tratamos de encontrar una curva de resonancia midiendo la absorción de luz infrarroja por el cloruro de sodio. Tal curva se muestra en la figura 23-7. La abscisa no es frecuencia, sino que está dada en términos de la longitud de onda, pero eso, por supuesto, es sólo un problema técnico, ya que para una onda existe una relación bien definida entre frecuencia y longitud de onda, de manera que es realmente una escala de frecuencia y una cierta frecuencia corresponde a la frecuencia de resonancia.

Pero ¿qué pasa con el ancho? ¿Qué determina el ancho? Hay muchos casos en los cuales el ancho que se ve en la curva no es realmente el ancho natural γ que uno obtendría teóricamente. Hay dos razones por las cuales puede haber una curva más ancha que la curva teórica. Si los objetos no tienen todos la misma frecuencia, como podría suceder si el cristal estuviera deformado en ciertas regiones de manera que en esas regiones la frecuencia de oscilación fuera ligeramente diferente que en otras regiones, entonces lo que tenemos son varias curvas de resonancia una encima de la otra, así que aparentemente obtenemos una curva más ancha. El otro tipo de ancho es simplemente éste, a lo mejor no podemos medir la frecuencia con suficiente precisión, si abrimos bastante la rendija del espectrómetro, entonces, aunque pensábamos que teníamos una sola frecuencia, realmente teníamos un cierto rango $\Delta\omega$, luego podríamos no tener el poder de resolución necesario para ver una curva angosta. De antemano, no podemos decir si el ancho de la figura 23-7 es natural, o si se debe a inhomogeneidades en el cristal o al ancho finito de la rendija en el espectrómetro.

Ahora cambiemos a un ejemplo más esotérico, el balanceo de un imán. Si tenemos un imán con polos norte y sur en un campo magnético constante, el extremo N del imán va a ser atraído hacia un lado y el extremo S hacia el otro, y en general actuará un torque sobre él de manera que va a oscilar en torno a su posición de equilibrio como la aguja de una brújula. Sin embargo, los imanes de que estamos hablando son *átomos*. Estos átomos tienen un momentum angular, el torque no produce por cierto un movimiento simple en la dirección del campo, sino una *precesión*. Ahora bien, mirado desde el lado, cualquier componente se está "balanceando" y podemos perturbar o inducir ese balanceo y medir una absorción. La curva en la figura 23-8 representa una curva de resonancia típica. Lo que se ha hecho aquí es ligeramente diferente desde el punto de vista técnico. La frecuencia del campo lateral que se usa para inducir este balanceo se mantiene siempre igual, mientras nosotros habíamos esperado que los investigadores la variaran e hicieran la curva. Ellos podrían haberlo hecho así, pero técnicamente para ellos fue más fácil dejar la frecuencia ω fija y variar la intensidad del campo magnético constante, lo que corresponde a variar ω_0 en nuestra fórmula. Dibujaron la curva de resonancia respecto a ω_0. De todos modos ésta es una resonancia típica con un cierto ω_0 y γ.

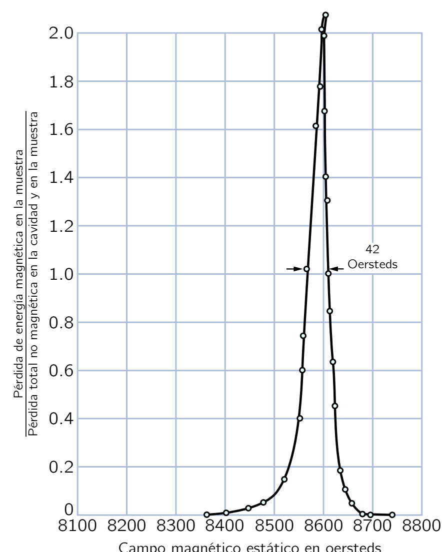


Figura 23-8. Pérdida de energía magnética en compuestos orgánicos paramagnéticos en función de la intensidad del campo magnético aplicado [Holden *et al.*, *Phys. Rev.* **75**, 1614 (1949)].

Vamos aún más allá. Nuestro próximo ejemplo tiene que ver con núcleos atómicos. Los movimientos de protones y neutrones en los núcleos son en cierto modo oscilatorios y podemos demostrar esto mediante el siguiente experimento. Bombardeamos un átomo de litio con protones y descubrimos que cierta reacción, que produce rayos γ, tiene en realidad un máximo muy agudo típico de resonancia. Notamos en la figura 23-10, sin embargo, una diferencia respecto a los otros casos, ¡la escala horizontal no representa una frecuencia, sino una *energía*! La razón es que en mecánica cuántica lo que consideramos clásicamente como la energía resulta estar realmente relacionada con una frecuencia de una amplitud de onda. Cuando analizamos algo que en la simple física a gran escala tiene que ver con una frecuencia, encontramos que cuando hacemos los experimentos cuánticos con materia atómica obtenemos la curva correspondiente en función de la energía. De hecho, en cierto sentido esta curva es una demostración de esta relación. Ella muestra que la frecuencia y la energía tienen cierta interrelación profunda, que por cierto poseen.

Cambiemos ahora a otro ejemplo que también incluye un nivel de energía nuclear, pero ahora uno mucho, mucho más estrecho. La ω_0 en la figura 23-9 corresponde a una energía de 100 000 electronvoltios, mientras que el ancho γ es aproximadamente 10^{-5} electronvoltios; en otras palabras, ¡esto tiene un Q de 10^{10}! Cuando se midió esta curva, fue el Q más grande de cualquier oscilador que se hubiera medido jamás. Fue medida por el doctor Mössbauer y fue la base de su premio Nobel. Aquí la escala horizontal es la velocidad, porque la técnica para obtener las frecuencias ligeramente diferentes consistió en usar el efecto Doppler moviendo la fuente respecto al absorbedor. Uno puede ver cuán delicado es el

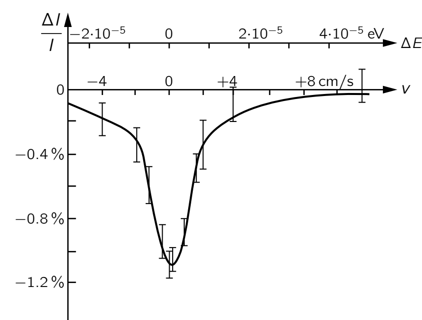


Figura 23-9. [Cortesía del doctor R. Mössbauer.]

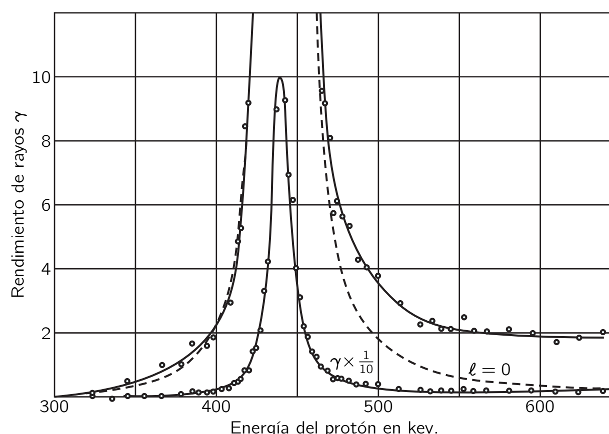


Figura 23-10. La intensidad de la radiación gamma del litio en función de la energía de los protones de bombardeo. La línea de puntos es teórica, calculada para protones con momentum angular $\ell = 0$. [Bonner y Evans, *Phys. Rev.* **73**, 666 (1948).]

experimento cuando nos damos cuenta de que ¡la velocidad en juego es de unos pocos centímetros por segundo! En la escala real de la figura la frecuencia cero correspondería a un punto a unos 10^{10} cm hacia la izquierda, ¡ligeramente fuera del papel!

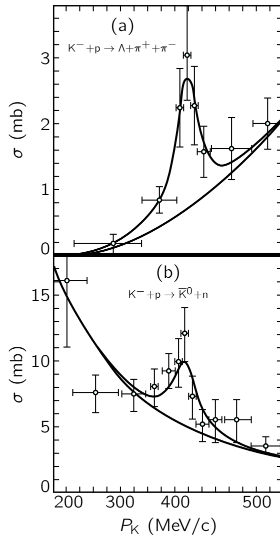


Figura 23-11. Dependencia del momentum en la sección eficaz para las reacciones (a) $K^- + p \rightarrow \Lambda + \pi^+ + \pi^-$ y (b) $K^- + p \rightarrow \overline{K}^0 + n$. Las curvas inferiores en (a) y (b) representan el fondo que se supone no es resonante, mientras que las de arriba contienen además la resonancia superpuesta. [Ferro-Luzzi *et al.*, *Phys. Rev. Lett*, **8**, 28 (1962).]

Finalmente, si examinamos un número del *Physical Review*, digamos el del 1° de enero de 1962, ¿encontraremos una curva de resonancia? Cada número tiene una curva de resonancia y la figura 23-11 es la curva de resonancia para éste. Esta curva de resonancia resulta muy interesante. Es la resonancia encontrada en una cierta reacción entre partículas extrañas, una reacción en la cual un K^- y un protón interactúan. La resonancia se detecta viendo cuántas partículas de algún tipo salen y dependiendo de cuáles y cuántas salen, uno obtiene diferentes curvas, pero de la misma forma y con el máximo agudo en la misma energía. Determinamos así que hay una resonancia a una cierta energía para el mesón K^-. Esto significa probablemente que existe algún tipo de estado o condición correspondiente a esta resonancia que se puede alcanzar poniendo juntos un K^- y un protón. Ésta es una nueva partícula o resonancia. Hoy día no sabemos si llamar un chichón como éste, una “partícula” o simplemente una resonancia. Cuando existe una resonancia muy *aguda*, corresponde a una *energía bien definida*, justamente como si hubiera una partícula con esa energía presente en la naturaleza. Cuando la resonancia se hace más ancha entonces no sabemos si decir que hay una partícula que no dura mucho o simplemente una resonancia en la probabilidad de reacción. En el segundo capítulo se indica esto para las partículas, pero cuando se escribió el segundo capítulo esta resonancia no era conocida ¡de manera que nuestra tabla debería contener aún otra partícula más!

24

Transitorios

24-1 La energía de un oscilador

Aunque este capítulo se titula "transitorios", algunas partes de él son, en cierto modo, parte del capítulo anterior sobre oscilaciones forzadas. Uno de los aspectos de una oscilación forzada que no hemos discutido todavía es la *energía* de la oscilación. Consideremos ahora esa energía.

¿Cuánta energía cinética hay en un oscilador mecánico? Es proporcional al cuadrado de la velocidad. Ahora llegamos a un punto importante. Consideren una cantidad arbitraria A que puede ser la velocidad o alguna otra cosa que queremos discutir. Cuando escribamos $A = \hat{A}e^{i\omega t}$, un número complejo, la A verdadera y honesta en el mundo físico es sólo la *parte real*; por lo tanto, si por alguna razón queremos usar el *cuadrado* de A no es correcto elevar al cuadrado el número complejo y luego tomar la parte real, porque la parte real del cuadrado de un número complejo no es simplemente el cuadrado de la parte real, sino que también incluye la parte *imaginaria*. De manera que cuando queremos encontrar la energía tenemos que apartarnos de la rotación compleja por un rato para ver cuáles son los funcionamientos internos.

La verdadera A física es la parte real de $A_0e^{i(\omega t+\Delta)}$, o sea, $A = A_0 \cos(\omega t+\Delta)$, donde $\hat{A}$, el número complejo, se escribe como $A_0e^{i\Delta}$. Ahora bien, el cuadrado de esta cantidad física real es $A^2 = A_0^2 \cos^2(\omega t + \Delta)$. El cuadrado de la cantidad va entonces hacia arriba y hacia abajo desde un máximo a cero, como el cuadrado del coseno. El cuadrado del coseno tiene un máximo de 1 y un mínimo de 0 y su valor medio es 1/2.

En muchos casos no estamos interesados en la energía en un momento específico durante la oscilación, para un gran número de aplicaciones sólo queremos el promedio de A^2, el *promedio* del cuadrado de A en un periodo de tiempo grande comparado con el periodo de oscilación. En estas circunstancias se puede usar el promedio del coseno cuadrado; de esta manera tenemos el siguiente teorema: si A está representado por un número complejo, entonces el promedio de A^2 es igual a $\frac{1}{2}A_0^2$. Ahora bien, A_0^2 es el cuadrado del módulo del complejo $\hat{A}$. (Esto se puede escribir de varias maneras, a algunas personas les gusta escribir $|\hat{A}|^2$, otras escriben $\hat{A}\hat{A}^*$, $\hat{A}$ es su complejo conjugado.) Vamos a usar este teorema varias veces.

Consideremos ahora la energía de un oscilador forzado. La ecuación para el oscilador forzado es

$$m\, d^2x/dt^2 + \gamma m\, dx/dt + m\omega_0^2 x = F(t). \tag{24.1}$$

En nuestro problema, por supuesto, $F(t)$ es una función coseno de t. Analicemos ahora la situación: ¿cuánto trabajo realiza la fuerza externa F? El trabajo realizado por la fuerza por segundo, en otras palabras, la potencia, es la fuerza por la velocidad. (Sabemos que el trabajo diferencial en el tiempo dt es $F dx$ y la potencia es $F dx/dt$.) Así,

$$P = F\,\frac{dx}{dt} = m\left[\left(\frac{dx}{dt}\right)\left(\frac{d^2x}{dt^2}\right) + \omega_0^2 x\left(\frac{dx}{dt}\right)\right] + \gamma m\left(\frac{dx}{dt}\right)^2. \tag{24.2}$$

Pero los dos primeros términos del segundo miembro también pueden escribirse $d/dt[\frac{1}{2}m(dx/dt)^2 + \frac{1}{2}m\omega_0^2x^2]$ que se puede verificar inmediatamente por diferenciación. Es decir, el término entre corchetes es una derivada pura de dos términos muy fáciles de comprender —uno es la energía cinética del movimiento y el

otro es la energía potencial del resorte—. Llamemos a esta cantidad la energía almacenada, es decir, la *energía almacenada* en la oscilación. Supongan que queremos la potencia media en muchos ciclos cuando el oscilador está siendo forzado y ha estado funcionando durante un tiempo largo. A la larga, la energía almacenada no cambia —su derivada da un efecto promedio cero—. En otras palabras, si promediamos la potencia en un tiempo largo, *toda la energía termina en última instancia en el término resistivo* $\gamma m(dx/dt)^2$. Hay *una cierta* energía almacenada en la oscilación, pero ésta no cambia con el tiempo si promediamos sobre muchos ciclos. Luego potencia media $\langle P\rangle$ es:

$$\langle P\rangle = \langle \gamma m(dx/dt)^2\rangle. \tag{24.3}$$

Usando nuestro método de escribir números complejos y nuestro teorema que $\langle A^2\rangle = 2A_0^2$ podemos encontrar esta potencia media. Así, si $x = \hat{x}e^{i\omega t}$, entonces $dx/dt = i\omega\hat{x}e^{i\omega t}$. Por lo tanto, en estas circunstancias, la potencia promedio se puede escribir como:

$$\langle P\rangle = \tfrac{1}{2}\gamma m\omega^2 x_0^2. \tag{24.4}$$

En la notación usada de circuitos eléctricos dx/dt se substituye por la corriente I (I es dq/dt donde q corresponde a x) y $m\gamma$ corresponde a la resistencia R. Así, la velocidad de pérdida de energía —la potencia consumida por la función forzante— es la resistencia en el circuito por el promedio de la corriente al cuadrado:

$$\langle P\rangle = R\langle I^2\rangle = R\cdot\tfrac{1}{2}I_0^2. \tag{24.5}$$

Esta energía, por supuesto, va a calentar la resistencia; se llama a veces pérdida de calor o calentamiento de Joule.

Otro aspecto interesante a discutir es cuánta energía está *almacenada.* Esto no es lo mismo que la potencia, porque aunque la potencia al principio se usó para almacenar algo de energía, después de eso el sistema continúa absorbiendo potencia mientras existan pérdidas por calentamiento (resistivas). En cualquier instante hay una cierta cantidad de energía almacenada, así que nos gustaría calcular también la energía media almacenada $\langle E\rangle$. Ya hemos calculado cuál es el promedio de $(dx/dt)^2$, así que encontramos:

$$\begin{aligned}\langle E\rangle &= \tfrac{1}{2}m\langle (dx/dt)^2\rangle + \tfrac{1}{2}m\omega_0^2\langle x^2\rangle\\ &= \tfrac{1}{2}m(\omega^2+\omega_0^2)\tfrac{1}{2}x_0^2.\end{aligned} \tag{24.6}$$

Ahora bien, cuando un oscilador es muy eficiente y ω es muy cercano a ω_0 de manera que $|\hat{x}|$ sea grande, la energía almacenada es muy alta —podemos obtener una energía almacenada grande a partir de una fuerza relativamente pequeña—. La fuerza realiza gran cantidad de trabajo para empezar la oscilación, pero para mantenerla constante todo lo que tiene que hacer es combatir la fricción. El oscilador puede tener una gran cantidad de energía si el roce es muy bajo y aunque esté oscilando intensamente no se pierde mucha energía. La eficiencia de un oscilador se puede medir por la cantidad de energía almacenada comparada con cuánto trabajo realiza la fuerza en cada oscilación.

¿Cómo se compara la energía almacenada con el trabajo realizado en un ciclo? Esto se llama el Q del sistema y Q se define como 2π por la energía almacenada promedio dividida entre el trabajo realizado por ciclo. (Si habláramos de trabajo realizado por *radián* en vez de por ciclo, desaparecería el 2π.)

$$Q = 2\pi\,\frac{\tfrac{1}{2}m(\omega^2+\omega_0^2)\cdot\langle x^2\rangle}{\gamma m\omega^2\langle x^2\rangle\cdot 2\pi/\omega} = \frac{\omega^2+\omega_0^2}{2\gamma\omega}. \tag{24.7}$$

Q no es un número muy útil, a menos que sea muy grande. Cuando es relativamente grande da una medida de lo bueno que es el oscilador. La gente ha tratado de definir Q de la manera más simple y útil; varias definiciones se diferencian un poco una de otra, pero si Q es muy grande todas las definiciones están de acuerdo. La definición más aceptada comúnmente es la ecuación (24.7) que depende de ω. Para un buen oscilador cerca de la resonancia, podemos simplificar un poco

(24.7) haciendo $\omega = \omega_0$ y entonces tenemos $Q = \omega_0/\gamma$ que es la definición de Q que usamos antes.

¿Qué es Q en un circuito eléctrico? Para averiguarlo basta que cambiemos m por L, $m\gamma$ por R y $m\omega_0^2$ por $1/C$ (véase tabla 23-1). La Q en la resonancia es $L\omega/R$, donde ω es la frecuencia de resonancia. Si consideramos un circuito con una Q grande, esto significa que la cantidad de energía almacenada en la oscilación es muy grande comparada con la cantidad de trabajo realizado por ciclo por la máquina que impulsa las oscilaciones.

24-2 Oscilaciones amortiguadas

Volvamos ahora a nuestro tema principal de discusión, los transitorios. Por *transitorio* se entiende una solución de la ecuación diferencial cuando no hay fuerza presente, pero el sistema no está simplemente en reposo. (Por supuesto, que si está en reposo en el origen sin que actúe ninguna fuerza, esto sí es un problema bonito —¡se queda ahí!—.) Supongan que la oscilación empieza de otra manera, digamos que fue impulsada por una fuerza durante un rato y después quitamos la fuerza. ¿Qué sucede entonces? Obtengamos primero una idea aproximada de lo que va a pasar en un sistema con una Q muy grande. Mientras la fuerza esté actuando, la energía almacenada permanecerá igual y se realiza una cierta cantidad de trabajo para mantenerla. Supongan ahora que quitamos la fuerza y no se realiza más trabajo, entonces las pérdidas que están consumiendo la energía de la fuente ya no las consumen más —ya no *hay* ningún impulsor—. Las pérdidas van a tener que consumir, por decirlo así, la energía que está almacenada. Supongan que $Q/2\pi = 1\,000$. Luego el trabajo realizado por ciclo es $1/1\,000$ de la energía almacenada. ¿No es razonable acaso, ya que está oscilando sin fuerza impulsora, que en un ciclo el sistema seguirá perdiendo un milésimo de su energía E que ordinariamente se habría estado suministrando desde el exterior y que continuará oscilando, siempre perdiendo $1/1\,000$ de su energía por ciclo? Así que, como hipótesis, para un sistema con una Q relativamente alta, supondremos que la siguiente ecuación podría ser aproximadamente válida (más adelante lo haremos exactamente ¡y va a resultar que *era* correcta!):

$$dE/dt = -\omega E/Q. \tag{24.8}$$

Esto es aproximado porque es válido sólo para Q grande. En cada radián, el sistema pierde una fracción $1/Q$ de la energía almacenada E. Luego, en una cantidad de tiempo dada dt la energía va a cambiar en una cantidad $\omega dt/Q$, ya que el número de radianes asociados con el tiempo dt es ωdt. ¿Cuál es la frecuencia? Supongamos que el sistema se mueve tan delicadamente, casi sin fuerza, que si lo soltamos va a oscilar esencialmente con la misma frecuencia por sí mismo. Vamos a estimar entonces que ω sea la frecuencia de resonancia ω_0. Deducimos así de la ecuación (24.8) que la energía almacenada va a variar según

$$E = E_0 e^{-\omega_0 t/Q} = E_0 e^{-\gamma t}. \tag{24.9}$$

Ésta sería la medida de la *energía* en cualquier instante. ¿Cuál sería la fórmula, aproximada, para la amplitud de la oscilación en función del tiempo? ¿La misma? ¡No! La cantidad de energía en un resorte, digamos, es proporcional al *cuadrado* del desplazamiento, la energía cinética es proporcional al *cuadrado* de la velocidad, de manera que la energía total es proporcional al *cuadrado* del desplazamiento. Luego el desplazamiento, la amplitud de la oscilación, va a disminuir la mitad de rápido debido al cuadrado. En otras palabras, intuimos que la solución para el movimiento amortiguado transitorio va a ser una oscilación de frecuencia cerca de la frecuencia de resonancia ω_0, en la cual la amplitud del movimiento sinusoidal va a disminuir según $e^{-\gamma t/2}$:

$$x = A_0 e^{-\gamma t/2} \cos \omega_0 t. \tag{24.10}$$

Esta ecuación y la figura 24-1 nos dan una idea de lo que debemos esperar; tratemos ahora de analizar el movimiento *exactamente* resolviendo la ecuación diferencial del movimiento mismo.

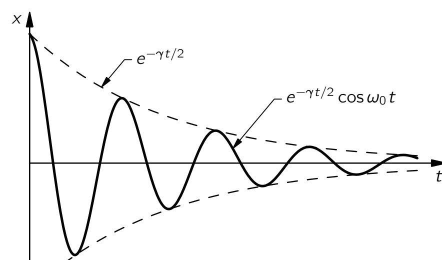


Figura 24-1. Una oscilación cosinusoidal amortiguada.

Así, partiendo de la ecuación (24.1) sin fuerza externa, ¿cómo la resolvemos? Siendo físicos no tenemos por qué preocuparnos tanto acerca del *método* como de cuál *es* la solución. Armados de nuestra experiencia previa, ensayemos cómo solucionar una curva exponencial $x = Ae^{i\alpha t}$. (¿Por qué ensayamos ésta? ¡Es la cosa más fácil de derivar!) La remplazamos en (24.1) (con $F(t) = 0$), usando la regla de que cada vez que derivemos x respecto al tiempo multiplicamos por $i\alpha$. Así que es realmente muy simple de sustituir. Luego nuestra ecuación tiene el siguiente aspecto:

$$(-\alpha^2 + i\gamma\alpha + \omega_0^2)Ae^{i\alpha t} = 0. \qquad (24.11)$$

El resultado neto debe ser cero para *todo tiempo*, lo que es imposible a menos que (a) $A = 0$, lo que no es ninguna solución, permanece quieto, o (b)

$$-\alpha^2 + i\alpha\gamma + \omega_0^2 = 0. \qquad (24.12)$$

Si podemos resolver esto y encontrar una α, ¡entonces tendremos una solución, en la cual A no necesita ser cero!

$$\alpha = i\gamma/2 \pm \sqrt{\omega_0^2 - \gamma^2/4}. \qquad (24.13)$$

Por un momento vamos a suponer que γ es bastante pequeño comparado con ω_0 de manera que $\omega_0^2 - \gamma^2/4$ es seguramente positivo y no hay ningún problema con extraer la raíz cuadrada. Lo único molesto es ¡que obtenemos *dos* soluciones! Así,

$$\alpha_1 = i\gamma/2 + \sqrt{\omega_0^2 - \gamma^2/4} = i\gamma/2 + \omega_\gamma \qquad (24.14)$$

y

$$\alpha_2 = i\gamma/2 - \sqrt{\omega_0^2 - \gamma^2/4} = i\gamma/2 - \omega_\gamma. \qquad (24.15)$$

Consideremos la primera, suponiendo que no nos hemos dado cuenta de que la raíz cuadrada tiene dos valores posibles. Entonces sabemos que una solución para x es $x_1 = Ae^{i\alpha_1 t}$, donde A es una constante cualquiera. Ahora, al sustituir α_1 llamaremos $\sqrt{\omega_0^2 - \gamma^2/4} = \omega_\gamma$, ya que va a aparecer tantas veces y demora tanto en escribirse. Así $i\alpha_1 = -\gamma/2 + i\omega_\gamma$ y obtenemos $x = Ae^{(-\gamma/2 + i\omega_\gamma)t}$, o lo que es lo mismo, debido a las propiedades maravillosas de una exponencial,

$$x_1 = Ae^{-\gamma t/2}e^{i\omega_\gamma t}. \qquad (24.16)$$

Primero, reconocemos esto como una oscilación, una oscilación a una frecuencia ω_γ que no es *exactamente* la frecuencia ω_0, pero está bastante cerca de ω_0 si es un buen sistema. Segundo, ¡la amplitud de la oscilación está decreciendo exponencialmente! Si tomamos, por ejemplo, la parte real de (24.16), obtenemos

$$x_1 = Ae^{-\gamma t/2}\cos\omega_\gamma t. \qquad (24.17)$$

Esto se parece mucho a la solución que intuimos (24.10), a excepción de la frecuencia que es realmente ω_γ. Éste es el único error, así que se trata de la misma cosa, tenemos la idea correcta. ¡Pero *no* todo está bien! Lo que está mal es que *hay otra solución.*

La otra solución es α_2 y vemos que la única diferencia es que el signo de ω_γ está cambiado:

$$x_2 = Be^{-\gamma t/2}e^{-i\omega_\gamma t}. \qquad (24.18)$$

¿Qué significa esto? Vamos a demostrar pronto que si x_1 y x_2 son soluciones posibles de la ecuación (24.1) con $F = 0$, ¡entonces $x_1 + x_2$ también es una solución de la misma ecuación! Así la solución general x tiene la forma matemática

$$x = e^{-\gamma t/2}(Ae^{i\omega_\gamma t} + Be^{-i\omega_\gamma t}). \qquad (24.19)$$

Ahora nos podemos preguntar por qué nos molestamos en dar esta otra solución, ya que estábamos tan contentos con la primera. ¿Para qué nos sirve la solución

extra ya que sabemos que debemos tomar sólo la parte real? *Sabemos* que debemos tomar la parte real, pero, ¿cómo iba a saber la *matemática* si sólo queríamos la parte real? Cuando teníamos una fuerza impulsora no nula $F(t)$, introdujimos una fuerza *artificial* junto con ella y la parte *imaginaria* de la ecuación, por decirlo así, era impulsada de una manera definida. Pero cuando hicimos $F(t) \equiv 0$, nuestra convención que x fuera sólo la parte real de todo lo que escribiéramos es sólo cosa nuestra, y las ecuaciones matemáticas no lo saben todavía. El mundo físico *tiene* una solución real, pero la respuesta con la que estábamos tan contentos antes no es real, es *compleja*. La *ecuación* no sabe que vamos a tomar arbitrariamente la parte real; así que tiene que presentársenos, por decirlo así, con una solución del tipo complejo conjugado de manera que al juntarlas podamos hacer una solución *verdaderamente real*; eso es lo que α_2 está haciendo por nosotros. Para que x sea real, $Be^{-\omega_\gamma t}$ va a tener que ser el complejo conjugado de $Ae^{i\omega_\gamma t}$ de manera que las partes imaginarias desaparezcan. Resulta así que B es el complejo conjugado de A y nuestra solución real es:

$$x = e^{-\gamma t/2}(Ae^{i\omega_\gamma t} + A^* e^{-i\omega_\gamma t}). \tag{24.20}$$

De modo que nuestra solución real es una oscilación con un *desfase* y una amortiguación, —como la anunciáramos—.

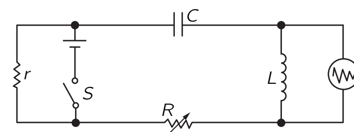


Figura 24-2. Un circuito eléctrico para demostrar transitorios.

24-3 Transitorios eléctricos

Veamos ahora si lo anterior realmente funciona. Construyamos el circuito eléctrico mostrado en la figura 24-2 en el cual aplicamos a un osciloscopio el voltaje a través de la inductancia L después de haber aplicado instantáneamente un voltaje cerrando el interruptor S. Es un circuito oscilatorio y genera un transitorio de algún tipo. Corresponde al caso en que aplicamos súbitamente una fuerza y el sistema empieza a oscilar. Es el análogo eléctrico de un oscilador mecánico amortiguado, y nosotros observamos la oscilación en un osciloscopio donde deberíamos observar las curvas que estamos tratando de analizar. (El movimiento horizontal del osciloscopio es a velocidad constante, mientras que el movimiento vertical es el voltaje a través del inductor. El resto del circuito es sólo un detalle técnico. Nos gustaría repetir el experimento muchas, muchas veces, ya que la persistencia de la visión no es suficientemente buena para ver sólo una traza en la pantalla. Así, pues, hacemos el experimento una y otra vez cerrando el interruptor 60 veces por segundo; cada vez que cerramos el interruptor empezamos el barrido horizontal del osciloscopio que dibuja la curva una y otra vez.) En las figuras 24-3 a 24-6 vemos ejemplos de oscilaciones amortiguadas, fotografiadas, realmente, de la pantalla de un osciloscopio. La figura 24-3 muestra una oscilación amortiguada en un circuito Q grande y γ chico. No muere muy rápido, oscila muchas veces al ir disminuyendo.

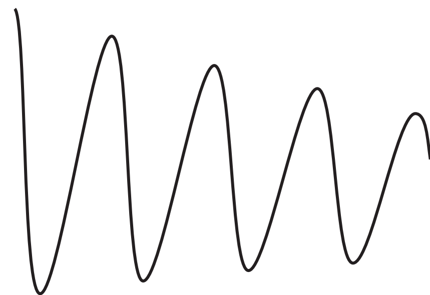


Figura 24-3

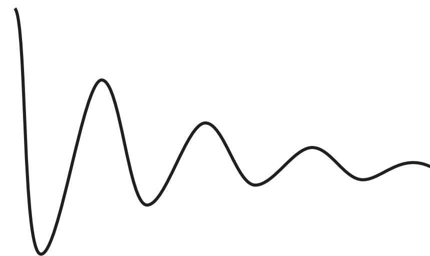


Figura 24-4

Pero veamos qué pasa cuando disminuimos Q, de manera que la oscilación se extinga más rápidamente. Podemos disminuir Q aumentando la resistencia R en el circuito. Cuando aumentamos la resistencia en el circuito, se extingue más rápidamente (figura 24-4). Luego, si aumentamos aún más la resistencia en el circuito, se extingue más rápidamente todavía (figura 24-5). Pero si vamos más allá de una cierta cantidad, ¡no podemos ver oscilación alguna! La pregunta es: ¿sucede esto porque nuestros ojos no son suficientemente buenos? Si aumentamos la resistencia aún más, obtenemos una curva como la de la figura 24-6, que no parece tener oscilaciones, excepto quizá una. ¿Cómo lo podemos explicar matemáticamente?

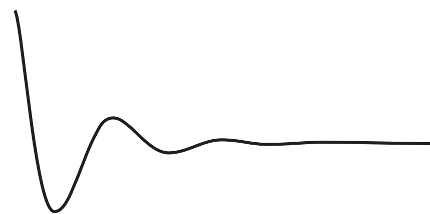


Figura 24-5

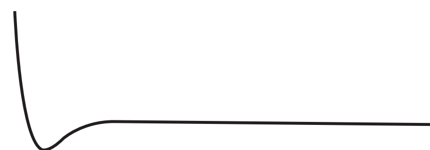


Figura 24-6

La resistencia es, por supuesto, proporcional al término γ en el dispositivo mecánico. Específicamente γ es R/L. Ahora bien, si aumentamos γ en las soluciones (24.14) y (24.15) con las que estábamos tan contentos, se establece un caos cuando $\gamma/2$ excede ω_0; tenemos que escribirlo de una manera diferente:

$$i\gamma/2 + i\sqrt{\gamma^2/4 - \omega_0^2} \qquad \text{e} \qquad i\gamma/2 - i\sqrt{\gamma^2/4 - \omega_0^2}.$$

Éstas son ahora las dos soluciones y siguiendo la misma línea de razonamiento matemático que antes, de nuevo encontramos dos soluciones: $e^{i\alpha_1 t}$ y $e^{i\alpha_2 t}$. Si ahora sustituimos α_1, obtenemos

$$x = Ae^{-(\gamma/2+\sqrt{\gamma^2/4-\omega_0^2})t},$$

una hermosa caída exponencial sin oscilación. Análogamente la otra solución es

$$x = Be^{-(\gamma/2-\sqrt{\gamma^2/4-\omega_0^2})t}.$$

Noten que la raíz cuadrada no puede ser mayor que $\gamma/2$, porque aun cuando $\omega_0 = 0$, un término es exactamente igual al otro. Pero ω_0^2 se resta de $\gamma^2/4$, de manera que la raíz cuadrada es menor que $\gamma/2$ y el término entre paréntesis es, por lo tanto, siempre un número positivo. ¡Gracias a Dios! ¿Por qué? Porque si fuera negativo, encontraríamos a e elevado a un factor *positivo* por t, ¡lo que significaría que estaba haciendo explosión! Al poner más y más resistencia en el circuito sabemos que no va a hacer explosión —muy por el contrario—. De manera que ahora tenemos dos soluciones, cada una por sí una exponencial decreciente, pero una con "velocidad de extinción" mucho mayor que la otra. La solución general es, por supuesto, una combinación de las dos; los coeficientes de la combinación dependen de cómo se empieza el movimiento —cuáles son las condiciones iniciales del problema—. En la forma particular según la cual este circuito se pone en funcionamiento, sucede que A es negativa y B es positiva; así que obtenemos la diferencia de dos curvas exponenciales.

Discutamos ahora cómo podemos encontrar los dos coeficientes A y B (o A y A^*) si sabemos cómo empezó el movimiento.

Supongan que para $t = 0$ sabemos que $x = x_0$ y $dx/dt = v_0$. Si ponemos $t = 0$, $x = x_0$ y $dx/dt = v_0$ en las expresiones:

$$x = e^{-\gamma t/2}(Ae^{i\omega_\gamma t} + A^* e^{-i\omega_\gamma t}),$$
$$dx/dt = e^{-\gamma t/2}[(-\gamma/2 + i\omega_\gamma)Ae^{i\omega_\gamma t} + (-\gamma/2 - i\omega_\gamma)A^* e^{-i\omega_\gamma t}],$$

encontramos, ya que $e^0 = e^{i0} = 1$,

$$\begin{aligned} x_0 &= A + A^* = 2A_R, \\ v_0 &= -(\gamma/2)(A + A^*) + i\omega_\gamma(A - A^*) \\ &= -\gamma x_0/2 + i\omega_\gamma(2iA_I), \end{aligned}$$

donde $A = A_R + iA_I$ y $A^* = A_R - iA_I$. Encontramos así:

$$A_R = x_0/2$$

y

$$A_I = -(v_0 + \gamma x_0/2)/2\omega_\gamma. \tag{24.21}$$

Esto determina completamente a A y A^* y por lo tanto la curva completa de la solución transitoria en términos de cómo comienza. A propósito, podemos escribir la solución de otra manera si notamos que:

$$e^{i\theta} + e^{-i\theta} = 2\cos\theta \qquad \text{y} \qquad e^{i\theta} - e^{-i\theta} = 2i \operatorname{sen}\theta.$$

Podemos entonces escribir la solución completa en la forma

$$x = e^{-\gamma t/2}\left[x_0 \cos\omega_\gamma t + \frac{v_0 + \gamma x_0/2}{\omega_\gamma} \operatorname{sen}\omega_\gamma t\right], \tag{24.22}$$

donde $\omega_\gamma = +\sqrt{\omega_0^2 - \gamma^2/4}$. Ésta es la expresión matemática para la forma en que una oscilación se extingue. No vamos a hacer uso directo de ella, pero hay una serie de puntos en los cuales nos gustaría poner énfasis, que son válidos en casos más generales.

Primero que nada, el comportamiento de un sistema tal sin fuerzas externas se expresa mediante una suma, o superposición, de exponenciales, puras en el tiempo

(que escribimos $e^{i\alpha t}$). Es una buena solución a probar en esas circunstancias. Los valores de α pueden ser complejos en general, representando las partes imaginarias el amortiguamiento. Finalmente, la íntima relación matemática de las funciones sinusoidales y exponenciales discutidas en el capítulo 22, aparecen muy a menudo físicamente como un cambio del comportamiento sinusoidal o exponencial cuando algún parámetro físico (en este caso la resistencia, γ) excede cierto valor crítico.

25

Sistemas lineales y repaso

25-1 Ecuaciones diferenciales lineales

En este capítulo discutiremos ciertos aspectos de los sistemas oscilantes, que se encuentran con una generalidad algo mayor que precisamente en los sistemas particulares que hemos estado estudiando. Para nuestro sistema particular, la ecuación diferencial que hemos estado resolviendo es:

$$m\,\frac{d^2x}{dt^2} + \gamma m\,\frac{dx}{dt} + m\omega_0^2 x = F(t). \tag{25.1}$$

Ahora bien, esta combinación particular de "operaciones" con la variable x tiene la interesante propiedad que, si sustituimos x por $(x+y)$, obtenemos la suma de las mismas operaciones con x y y; o si multiplicamos x por a entonces obtenemos a por la misma combinación. Esto es fácil de demostrar. Al igual que en una notación "taquigráfica", como nos cansamos de escribir todas esas letras en (25.1), usaremos el símbolo $\underline{L}(x)$ en su lugar. Cuando veamos esto, significa el primer miembro de (25.1) en la cual ha sido sustituido x. Con esta manera de escribir, $\underline{L}(x+y)$ significaría lo siguiente:

$$\underline{L}(x+y) = m\,\frac{d^2(x+y)}{dt^2} + \gamma m\,\frac{d(x+y)}{dt} + m\omega_0^2(x+y). \tag{25.2}$$

(Subrayamos la $\underline{L}$ para acordarnos de que no es una función ordinaria.) A veces llamamos a esto *notación operacional* pero no tiene importancia cómo lo llamemos, es "taquigrafía" simplemente.

Nuestra primera afirmación fue que:

$$\underline{L}(x+y) = \underline{L}(x) + \underline{L}(y), \tag{25.3}$$

que por supuesto se deduce del hecho que $a(x+y) = ax + ay$, $d(x+y)/dt = dx/dt + dy/dt$, etcétera.

Nuestra segunda afirmación fue, para a constante,

$$\underline{L}(ax) = a\underline{L}(x). \tag{25.4}$$

(Realmente (25.3) y (25.4) están muy relacionadas porque si ponemos $x+x$ en (25.3), es lo mismo que poner $a=2$ en (25.4), etcétera).

En problemas más complicados puede haber más derivadas y más términos en $\underline{L}$, lo interesante es si se mantienen o no las dos ecuaciones (25.3) y (25.4). Si lo hacen, a tal problema lo llamamos problema *lineal.* En este capítulo vamos a discutir algunas propiedades que existen debidas al hecho que el sistema es lineal, para apreciar la generalidad de algunos de los resultados que obtuvimos en nuestro análisis especial de una ecuación especial.

Estudiemos ahora algunas propiedades de las ecuaciones diferenciales lineales habiéndolas ya ilustrado con la ecuación específica (25.1) que estudiamos tan exhaustivamente. La primera propiedad de interés es la siguiente: supongan que tenemos que resolver la ecuación diferencial de un transitorio, la oscilación libre sin fuerza impulsora. Es decir, queremos resolver:

$$\underline{L}(x) = 0. \tag{25.5}$$

Supongan que por alguna artimaña hemos encontrado una solución particular, que llamaremos x_1. Es decir, tenemos una x_1 para la cual $\underline{L}(x_1) = 0$. Ahora notamos que ax_1 también es una solución de la misma ecuación, podemos

multiplicar esta solución especial por cualquier constante y obtener una nueva solución. En otras palabras, si tuviéramos un movimiento de un cierto "tamaño", un movimiento el doble de "grande" también es una solución. *Demostración*: $\underline{L}(ax_1) = a\underline{L}(x_1) = a \cdot 0 = 0$.

Enseguida supongan que por otra artimaña hemos encontrado no sólo *una* solución x_1, sino que también otra solución x_2. (Recuerden que cuando sustituimos $x = e^{i\alpha t}$ para encontrar los fenómenos transitorios encontramos *dos* valores de α, es decir, dos soluciones x_1 y x_2.) Mostremos ahora que la combinación $(x_1 + x_2)$ también es una solución. En otras palabras, si hacemos $x = x_1 + x_2$, x también es una solución de la ecuación. ¿Por qué? Porque, si $\underline{L}(x_1) = 0$ y $\underline{L}(x_2) = 0$, entonces $\underline{L}(x_1 + x_2) = \underline{L}(x_1) + \underline{L}(x_2) = 0 + 0 = 0$. Luego, si hemos encontrado un cierto número de soluciones para el movimiento de un sistema lineal, las podemos sumar.

Combinando estas dos ideas, vemos por supuesto que podemos también sumar seis de una y dos de otra, si x_1 es una solución también lo es αx_1. Por lo tanto, cualquier suma de estas dos soluciones, como $(\alpha x_1 + \beta x_2)$, es también una solución. Si sucede que podemos encontrar tres soluciones hallamos entonces que cualquier combinación de estas tres soluciones es también una solución, etc. Resulta que el número de lo que llamamos *soluciones independientes** que hemos obtenido para nuestro problema del oscilador es solamente *dos*. El número de soluciones independientes que uno encuentra en el caso general depende de lo que se llama el número de *grados de libertad*. No vamos a discutir esto en detalle ahora, pero si tenemos una ecuación diferencial de segundo orden, hay sólo soluciones independientes y hemos encontrado ambas; tenemos así la solución más general.

Vamos ahora a otra proposición que se aplica a la situación en la cual el sistema está sometido a una fuerza externa. Supongan que la ecuación es:

$$\underline{L}(x) = F(t), \tag{25.6}$$

y supongan que le hemos encontrado una solución especial. Digamos que la solución de Juan es x_J y que $\underline{L}(x_J) = F(t)$. Supongan que queremos encontrar otra solución, supongan que sumamos a la solución de Juan una de las que era una solución de la ecuación libre (25.5), digamos x_1. Entonces vemos, según (25.3), que:

$$\underline{L}(x_J + x_1) = \underline{L}(x_J) + \underline{L}(x_1) = F(t) + 0 = F(t). \tag{25.7}$$

Por lo tanto, a la solución "forzada" le podemos sumar cualquier solución "libre" y aún tenemos una solución. La solución "libre" se llama solución *transitoria*.

Cuando no tenemos ninguna fuerza actuando y súbitamente ponemos una, no obtenemos inmediatamente la solución estacionaria que resolvimos con la solución sinusoidal, sino que por un tiempo hay un transitorio que tarde o temprano se extingue si esperamos lo suficiente. La solución "forzada" no se extingue ya que sigue siendo impulsada por la fuerza. En último término, para periodos grandes de tiempo la solución es única, pero inicialmente los movimientos son diferentes por circunstancias distintas, dependiendo de la manera como empezó el sistema.

25-2 Superposición de soluciones

Ahora nos encontramos con otra proposición interesante. Supongan que tenemos una cierta fuerza impulsora particular F_a (digamos que es oscilatoria, con un cierto $\omega = \omega_a$ aunque nuestras conclusiones van a ser válidas para cualquier forma funcional de F_a) y que hemos encontrado la solución para el movimiento forzado (con o sin los transitorios, no tiene importancia). Supongan ahora que actúa alguna otra fuerza, digamos F_b, y que resolvemos el mismo problema pero con fuerza diferente. Entonces supongan que viene alguien y dice "tengo un nuevo problema para que ustedes lo resuelvan, tengo la fuerza $F_a = F_b$". ¿Podemos hacerlo? Por supuesto que lo podemos hacer, porque la solución es la suma

* Las soluciones que no pueden ser expresadas como combinaciones lineales, las unas de las otras, se llaman independientes.

de las dos soluciones x_a y x_b para las fuerzas tomadas separadamente —una circunstancia por cierto muy notable—. Si usamos (25.3), vemos que:

$$\underline{L}(x_a + x_b) = \underline{L}(x_a) + \underline{L}(x_b) = F_a(t) + F_b(t). \tag{25.8}$$

Esto es un ejemplo de lo que se llama *principio de superposición* para sistemas lineales, y es muy importante. Significa lo siguiente: si tenemos una fuerza complicada que se puede descomponer de cualquier manera conveniente en la suma de partes separadas, cada una de las cuales es en cierto modo simple en el sentido de que para cada parte especial en que hemos dividido la fuerza podemos resolver la ecuación, entonces el resultado es válido para la fuerza *total*, porque simplemente podemos volver a juntar las partes de la *solución* de la misma manera como la *fuerza* total está formada de partes (figura 25-1).

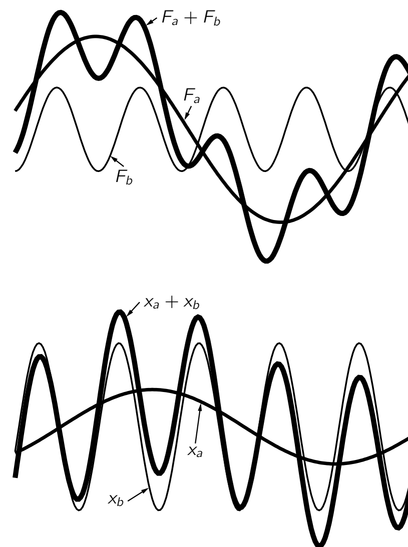


Figura 25-1. Un ejemplo del principio de superposición para sistemas lineales.

Demos otro ejemplo del principio de superposición. En el capítulo 12 dijimos que uno de los grandes hechos de las leyes de la electricidad era que si tenemos una cierta distribución de cargas q_a y calculamos el campo eléctrico $\boldsymbol{E}_a$ producidas por estas cargas en un cierto lugar P y si, por otro lado, tenemos otro conjunto de cargas q_b y calculamos el campo F_b debido a éstas en el lugar correspondiente, entonces, si ambas distribuciones de cargas están presentes al mismo tiempo, el campo $\boldsymbol{E}$ en P es la *suma* de $\boldsymbol{E}_a$ debido a un conjunto más $\boldsymbol{E}_b$ debido al otro. En otras palabras, si conocemos el campo debido a cierta carga, el campo debido a muchas cargas es simplemente el vector suma de los campos de estas cargas tomados individualmente. Esto es exactamente análogo a la proposición anterior que si conocemos el resultado de dos fuerzas dadas tomadas una a una; entonces, si la fuerza se considera como la suma de ellas, la respuesta es una suma de las respuestas individuales correspondientes.

La razón de que esto sea válido en la electricidad es que las grandes leyes de la electricidad, las ecuaciones de Maxwell, que determinan el campo eléctrico, son ecuaciones diferenciales *lineales*, es decir, que tienen la propiedad (25.3). Lo que corresponde a la fuerza es la *carga* que genera el campo eléctrico y la ecuación que determina el campo eléctrico en términos de la carga es lineal.

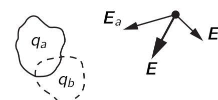


Figura 25-2. El principio de superposición en electrostática.

Como otro ejemplo interesante de esta proposición, preguntamos cómo es posible "sintonizar" cierta estación de radio al mismo tiempo que todas las otras emisoras están transmitiendo. La estación de radio transmite, fundamentalmente, un campo eléctrico oscilante de muy alta frecuencia que actúa sobre nuestra antena de radio. Es cierto que la amplitud de oscilación del campo es modificada, modulada, para transportar la señal de la voz, pero eso es muy lento y no nos vamos a preocupar por ello. Cuando uno escucha "Esta estación está transmitiendo con una frecuencia de 780 kilociclos", ello indica que 780 000 oscilaciones por segundo es la frecuencia del campo eléctrico de la antena de la estación y esto arrastra en nuestra antena los electrones hacia arriba y hacia abajo con esa frecuencia. Podemos tener ahora al mismo tiempo otra estación de radio en el mismo pueblo radiando con una frecuencia diferente, digamos 550 kilociclos por segundo, luego, los electrones en nuestra antena también están siendo arrastrados a esa frecuencia. La pregunta ahora es, ¿cómo podemos separar las señales que llegan a una radio a 780 kilociclos de las que llegan a 550 kilociclos? Ciertamente no escuchamos ambas estaciones al mismo tiempo.

Por el principio de superposición, la respuesta del circuito eléctrico de la radio, la primera parte del cual es un circuito lineal, a las fuerzas que están actuando debido al campo eléctrico $F_a + F_b$, es $x_a + x_b$. Pareciera entonces como que nunca las vamos a desenredar. De hecho, el mismo enunciado de superposición parece insistir en que no podemos *evitar* tener a ambas en nuestro sistema. Pero recuerden, para un circuito *resonante*, la curva respuesta, la cantidad de x por unidad de F en función de la frecuencia, se ve como la de la figura 25-3. Si fuera un circuito con un Q muy alto, la respuesta mostraría un máximo muy agudo. Supongan ahora que las dos estaciones son comparables en intensidad, o sea, que las dos *fuerzas* sean de la misma magnitud. La *respuesta* que obtenemos es la suma de x_a y x_b. Pero en la figura 25-3, x_a es tremenda mientras que x_b es pequeña. Así, a pesar de que las dos señales tienen la misma intensidad,

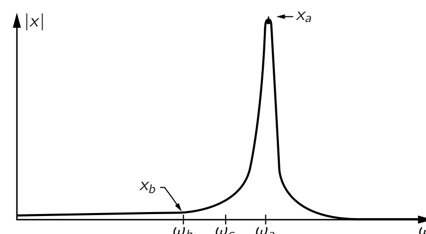


Figura 25-3. Una curva de resonancia de sintonía aguda.

cuando pasan por el circuito resonante agudo de la radio sintonizada para ω_a, la frecuencia de transmisión de una estación, entonces la respuesta a esta estación es mucho mayor que a la otra. Por lo tanto, la respuesta completa con las dos señales está casi toda formada por ω_a y hemos seleccionado la emisora que queríamos.

¿Y qué pasa con la sintonía? ¿Cómo sintonizamos? Cambiamos ω_0 cambiando la L o la C del circuito porque la frecuencia del circuito tiene que ver con la combinación de L y C. En particular, la mayoría de las radios están construidas de modo que uno pueda cambiar la capacitancia. Cuando se vuelva a sintonizar la radio podemos cambiar la posición del dial de manera que la frecuencia natural del circuito se desplaza, digamos, a ω_c. En este caso no escuchamos ni una ni otra emisora; obtenemos silencio siempre que no haya otra emisora con frecuencia ω_c. Si seguimos cambiando la capacitancia hasta que la curva de resonancia esté en ω_b, entonces por cierto escuchamos la otra emisora. Así es como funciona la sintonía de una radio, es de nuevo el principio de superposición combinado con una respuesta resonante.*

Para terminar esta discusión, describamos cualitativamente lo que sucede si proseguimos analizando un problema lineal con una fuerza dada cuando la fuerza es bastante complicada. De los muchos métodos posibles hay dos maneras generales especialmente útiles según las cuales podemos resolver el problema. Una es ésta: supongan que podemos resolverlo para fuerzas especiales conocidas como sinusoides de diferentes frecuencias. Sabemos que es juego de niños resolverlo para sinusoides. Así tenemos los que llamaremos casos de “juego de niño”. Ahora bien, la pregunta es si nuestra fuerza tan complicada se puede representar como la suma de dos o más fuerzas de “juego de niños”. En la figura 25-1 ya teníamos una curva bastante complicada y, por supuesto, la podemos hacer aún más complicada si agregamos más sinusoides. De manera que es ciertamente posible obtener curvas muy complicadas. Y, de hecho, lo inverso también es cierto: se puede obtener prácticamente cualquier curva sumando un *número infinito* de sinusoides de diferentes longitudes de onda (o frecuencias) para cada una de las cuales conocemos el resultado. Simplemente tenemos que conocer cuánto poner de cada sinusoide para formar la F dada, y entonces nuestra respuesta x es la suma correspondiente de las F sinusoides, cada una multiplicada por su respectiva razón de x a F. Este método de solución se llama método de la *transformada de Fourier* o *análisis de Fourier*. Realmente no vamos a realizar este análisis ahora, sólo deseamos describir la idea en juego.

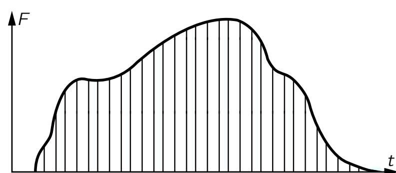


Figura 25-4. Una fuerza complicada se puede tratar como una sucesión de impulsos agudos.

Otra manera muy interesante de resolver nuestro problema tan complicado es la siguiente. Supongan que mediante un esfuerzo mental tremendo fuera posible resolver nuestro problema para una fuerza especial, a saber, un *impulso*. La fuerza se aplica rápidamente y enseguida se elimina, desaparece completamente. En realidad necesitamos resolver sólo para un impulso de cierta intensidad unitaria, cualquier otra intensidad se puede obtener multiplicando por un factor apropiado. Sabemos que la respuesta x a un impulso es una oscilación amortiguada. ¿Qué podemos decir de alguna otra fuerza, por ejemplo una fuerza como la de la figura 25-4?

Tal fuerza puede asemejarse a una sucesión de golpes con un martillo. Primero no hay fuerza y de súbito hay una fuerza constante —impulso, impulso, impulso, impulso. . . y luego cesa—. En otras palabras, imaginamos que la fuerza continua es una serie de impulsos muy juntos. Sabemos el resultado para un impulso, de manera que el resultado para toda una serie de impulsos va a ser toda una serie de oscilaciones amortiguadas, va a ser la curva para el primer impulso y luego (un poquito después) le sumamos la curva para el segundo impulso y la curva para el tercer impulso, y así sucesivamente. Por lo tanto, podemos representar matemáticamente la solución completa para funciones arbitrarias, si conocemos

* En los receptores modernos superheterodinos la operación real es más compleja. Los amplificadores están todos sintonizados a una frecuencia fija (llamada frecuencia FI) y un oscilador de frecuencia sintonizable variable se combina con la señal de entrada en un circuito no lineal para producir una nueva frecuencia (la diferencia de las frecuencias de la señal y del oscilador) igual a la frecuencia FI, que entonces es amplificado. Esto va a ser tratado en el capítulo 50.

la respuesta para un impulso. Obtenemos la respuesta para cualquier otra fuerza simplemente por integración. Este método se llama *método de la función de Green.* Una función de Green es una respuesta a un impulso y el método para analizar cualquier fuerza juntando las respuestas de impulsos se llama método de la función de Green.

Los principios físicos encerrados en estos dos esquemas son tan simples, incluyendo sólo la ecuación lineal, que se pueden comprender inmediatamente, pero los problemas *matemáticos* que entran en juego, las integraciones complicadas, etc., son demasiado avanzadas para atacarlas ahora. Con toda seguridad ustedes van a volver sobre esto algún día, cuando hayan tenido más práctica en matemática. Pero la *idea* es en realidad muy simple.

Finalmente, hagamos algunas observaciones de por qué los sistemas *lineales* son tan importantes. La respuesta es simple, ¡porque los podemos resolver! De manera que la mayor parte del tiempo resolvemos problemas lineales. Segundo (y más importante), resulta que las *leyes fundamentales de la física son a menudo lineales.* Por ejemplo, las ecuaciones de Maxwell para las leyes de la electricidad son lineales. Las grandes leyes de la mecánica cuántica resultan ser hasta donde sabemos ecuaciones lineales. Por *eso* es que gastamos tanto tiempo en ecuaciones lineales, porque si entendemos las ecuaciones lineales, estamos listos, en principio, para entender una cantidad de otras cosas.

Mencionemos otra situación donde se encuentran ecuaciones lineales. Cuando los desplazamientos son pequeños, muchas funciones pueden ser *aproximadas* linealmente. Por ejemplo, si tenemos un péndulo simple, la ecuación correcta para su movimiento es:

$$d^2\theta/dt^2 = -(g/L)\,\mathrm{sen}\,\theta. \tag{25.9}$$

Esta ecuación se puede resolver mediante funciones elípticas, pero la manera más fácil de resolverla es numéricamente como se demostró en el capítulo 9 en las leyes de movimiento de Newton. Comúnmente, una ecuación no lineal no se puede resolver más que en forma numérica. Ahora bien, para θ pequeño, $\mathrm{sen}\,\theta$ es prácticamente igual a θ y tenemos una ecuación lineal. Resulta que hay muchos casos donde pequeños efectos son lineales, como este ejemplo, el balanceo de un péndulo en pequeños arcos. Como otro ejemplo, si tiramos un poco de un resorte, la fuerza es proporcional al estiramiento. Si tiramos muy fuerte, rompemos el resorte y ¡la fuerza es una función completamente diferente de la distancia! Las ecuaciones lineales son importantes. En realidad son *tan* importantes que a lo mejor 50 % estamos resolviendo ecuaciones lineales en física e ingeniería.

25-3 Oscilaciones en sistemas lineales

Repasemos ahora las cosas sobre las cuales hemos estado hablando en los últimos capítulos. Es muy fácil que la física de los osciladores se haga oscura por la matemática. La física es realmente muy simple y si pudiéramos olvidarnos de la matemática por un momento veríamos que podemos entender casi todo lo que sucede en un sistema oscilante. Primero, si tenemos sólo el resorte y el peso, es fácil de entender por qué el sistema oscila —es una consecuencia de la inercia—. Tiramos la masa hacia abajo y la fuerza la tira hacia arriba, cuando pasa por cero, que es el lugar en que le gusta estar, no puede detenerse súbitamente, debido a su momentum sigue moviéndose y oscila hacia el otro lado y continúa de un lado a otro. De manera que si no hubiera fricción, seguramente esperaríamos un movimiento oscilatorio y efectivamente obtenemos uno. Pero si hay un poco de fricción siquiera, entonces en el ciclo de vuelta la oscilación no va a ser tan grande como fue la primera vez.

¿Qué es lo que pasa ahora, ciclo a ciclo? Eso depende del tipo y de la cantidad de fricción. Supongan que pudiéramos confeccionar un tipo de fuerza de fricción que siempre se mantiene en la misma proporción respecto a las otras fuerzas, la de inercia y la del resorte, a medida que la amplitud de la oscilación varía. En otras palabras, para oscilaciones pequeñas la fricción debería ser más débil

que para oscilaciones grandes. La fricción ordinaria no tiene esta propiedad, de manera que se debe inventar un tipo especial de fricción cuidadosamente con el especial propósito de crear una fricción que es directamente proporcional a la velocidad, de manera que para grandes oscilaciones sea mayor y para pequeñas oscilaciones sea más débil. Si sucede que tenemos ese tipo de fricción, al final de cada ciclo sucesivo el sistema está en la misma condición en que estaba al principio, excepto que es un poco más chico. Todas las fuerzas son más chicas en la misma proporción, la fuerza del resorte se reduce, los efectos inerciales son menores porque las aceleraciones son ahora más débiles y la fricción es menor también porque así lo concebimos cuidadosamente. Cuando realmente tenemos ese tipo de fricción, encontramos que cada oscilación es exactamente la misma que la primera, excepto que está reducida en amplitud. Si el primer ciclo disminuyó la amplitud a, digamos, 90 % de lo que era al principio, la siguiente la va a disminuir a 90 % de 90 %, etc., *el tamaño de las oscilaciones se reduce en la misma fracción en cada ciclo*. Una función exponencial es una curva que hace exactamente eso. Cambia en el mismo factor en cada intervalo igual de tiempo. Es decir, si la amplitud de un ciclo relativa a la anterior se llama a, entonces la amplitud de la siguiente es a^2 y de la siguiente a^3. De manera que la amplitud es alguna constante elevada a una potencia igual al número de ciclos recorridos:

$$A = A_0 a^n. \tag{25.10}$$

Pero, por supuesto, $n \propto t$, así que está perfectamente claro que la solución general va a ser algún tipo de oscilación, seno o coseno ωt, por una amplitud que varía más o menos como b^t. Pero b se puede escribir como e^{-c} si b es positivo y menor que 1. Por esto es que la solución aparece como $e^{-ct} \cos \omega t$. Es muy simple.

¿Qué sucede si la fricción no es tan artificial, por ejemplo, la fricción común de una mesa, de manera que la fuerza de fricción es una cierta cantidad constante e independiente de la magnitud de la oscilación que invierte su dirección cada medio ciclo? Entonces la ecuación ya no es más lineal, se hace difícil de resolver y debe ser resuelta por el método numérico dado en el capítulo 2, o considerando cada medio ciclo separadamente. El método numérico es el más poderoso de todos y puede resolver cualquier ecuación. Solamente cuando tenemos un problema simple podemos usar el análisis matemático.

El análisis matemático no es la cosa grandiosa que se dice ser, resuelve sólo las ecuaciones más simples. Tan pronto como las ecuaciones se hacen un poco más complicadas, nada más que un poquito —no se las puede resolver analíticamente—. Pero el método numérico, al cual se le hizo propaganda al principio del curso, puede hacerse cargo de cualquier ecuación de interés físico.

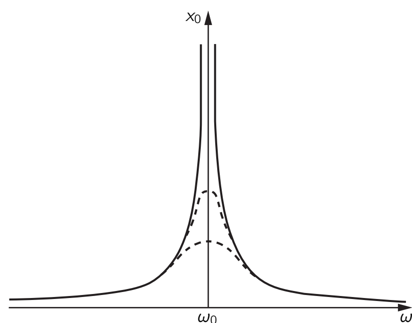


Figura 25-5. Curvas de resonancia en presencia de diversas cantidades de fricción.

A continuación, ¿qué pasa con la curva de resonancia? ¿Dónde está la resonancia? Primero, imaginen por un momento que no hay fricción y tenemos algo que pueda oscilar por sí mismo. Si golpeáramos el péndulo en forma precisa cada vez que pasa, por supuesto que podríamos hacerlo mover como loco. Pero si cerramos los ojos y no lo miramos y lo golpeamos a intervalos iguales arbitrarios, ¿qué va a suceder? A veces nos vamos a encontrar golpeando cuando va en el sentido contrario. Cuando sucede que tenemos el ritmo correcto, por supuesto, cada golpe va a ser dado justo a tiempo de manera que va cada vez más alto y más alto y más alto. De manera que sin fricción obtendríamos una curva como la curva llena de la figura 25-5 para diferentes frecuencias. Cualitativamente, comprendemos la curva de resonancia; con el fin de obtener la forma exacta de la curva probablemente lo más apropiado sea hacerlo matemáticamente. La curva tiende a infinito cuando $\omega \to \omega_0$ donde ω_0 es la frecuencia natural del oscilador.

Supongamos ahora que hay un poco de fricción; cuando el desplazamiento del oscilador es pequeño, la fricción no lo afecta mucho, la curva de resonancia es la misma, excepto cuando estamos cerca de la resonancia. En vez de hacerse infinita cerca de la resonancia, la curva va a alcanzar solamente una altura tal que el trabajo realizado por nuestros golpes cada vez es suficiente para compensar la pérdida de energía por fricción durante el ciclo. De manera que el ápice de la curva es redondeada —no va a infinito—. Si hay más fricción, el ápice de

la curva se redondea aún más. Ahora bien, alguien podría decir: "yo creía que los anchos de las curvas dependían de la fricción". Eso es porque la curva se traza comúnmente de manera que su ápice se representa por una unidad. Sin embargo, la expresión matemática es aún más simple de entender, si trazamos todas las curvas con la misma escala, entonces ¡todo lo que sucede es que la fricción rebaja el ápice! Si hay menos fricción, podemos subir más por ese pico antes que la fricción lo corte, de manera que se ve relativamente más angosto. Esto es, mientras más alto es el máximo de la curva, tanto más angosto es el ancho en la mitad de la altura máxima.

Finalmente, tomemos el caso de que haya una enorme cantidad de fricción. Resulta que si hay demasiada fricción, el sistema no oscila en absoluto. La energía del resorte es apenas capaz de moverlo en contra de las fuerzas de fricción, y así, se desliza lentamente hasta el punto de equilibrio.

25-4 Analogías en física

El siguiente aspecto a señalar en este repaso es que masas y resortes no son los únicos sistemas lineales, hay otros. En particular, hay sistemas eléctricos llamados circuitos lineales, en los cuales encontramos una analogía completa con los sistemas mecánicos. No hemos aprendido exactamente *por qué* cada uno de los objetos en un circuito eléctrico funciona en la forma que lo hace —no tiene por qué entenderlo en este momento—, podemos establecerlo como un hecho verificable experimentalmente que se comportan como se ha dicho.

Por ejemplo, tomemos el caso más simple posible. Tenemos un pedazo de alambre, que es una resistencia, y hemos aplicado a él una diferencia de potencial V. Ahora bien, V tiene este significado: si llevamos una carga q a lo largo del alambre desde un terminal al otro terminal, el trabajo realizado es qV. Cuanto más alta la diferencia de voltaje tanto más trabajo se realiza cuando la carga, como decimos, "cae" del extremo del terminal de potencial alto al extremo de potencial bajo. De manera que las cargas entregan energía al ir de un extremo al otro. Ahora bien, las cargas no vuelan simplemente derecho desde un extremo hasta el otro, los átomos del alambre ofrecen alguna resistencia a la corriente y esta resistencia obedece la siguiente ley para casi todas las sustancias ordinarias, si hay una corriente I, o sea, tantas cargas por segundo tropezando, el número por segundo que atraviesa tropezando el alambre es proporcional a la fuerza que las empujan; en otras palabras, proporcional a cuánto voltaje existe:

$$V = IR = R(dq/dt). \tag{25.11}$$

El coeficiente R se llama *resistencia* y la ecuación se llama ley de Ohm. La unidad de resistencia es el ohm, es igual a un voltio por amperio. En situaciones mecánicas es difícil obtener tal fuerza de fricción proporcional a la velocidad; en un sistema eléctrico es muy fácil y esta ley es extremadamente precisa para casi todos los metales.

Estamos a menudo interesados en cuánto trabajo se realiza por segundo, la pérdida de potencia o la energía liberada por las cargas a medida que tropiezan a lo largo del cable. Cuando llevamos una carga q a través de un voltaje V, el trabajo es qV, de manera que el trabajo realizado por segundo sería $V(dq/dt)$, que es lo mismo que VI o también $IR \cdot I = I^2R$. Esto se llama *pérdida por calentamiento*, es cuánto calor se genera en la resistencia por segundo, debido a la conservación de la energía. Es este calor el que hace que funcione una lámpara incandescente ordinaria.

Por supuesto, hay otras propiedades interesantes de sistemas mecánicos, como la masa (inercia), y resulta que hay también una analogía eléctrica para la inercia. Es posible hacer algo llamado *inductor* que tiene una propiedad llamada *inductancia* tal que una corriente, una vez establecida en la inductancia, *no quiere detenerse*. ¡Se necesita un voltaje para cambiar la corriente! Si la corriente es constante no hay voltaje a través de una inductancia. Los circuitos CC no saben nada de inductancia, es sólo cuando *cambiamos* la corriente que se manifiestan los efectos de la inductancia. La ecuación es:

$$V = L(dI/dt) = L(d^2q/dt^2), \tag{25.12}$$

y la unidad de inductancia, llamada *henrio*, es tal que un voltio aplicado a una inductancia de un henrio produce un cambio de un amperio por segundo en la corriente. La ecuación (25.12) es la analogía de la ley de Newton para la electricidad, si así lo deseamos, ¡V corresponde a F, L corresponde a m, e I corresponde a la velocidad! Todas las ecuaciones siguientes para los dos tipos de sistemas van a tener la misma deducción, porque en todas las ecuaciones podemos cambiar cualquier letra por su letra análoga correspondiente y obtenemos la misma ecuación, cualquier cosa que deduzcamos va a tener una correspondencia en los dos sistemas.

Ahora bien, ¿qué cosa eléctrica corresponde al resorte mecánico, donde existía una fuerza proporcional al estiramiento? Si empezamos con $F = kx$ y remplazamos $F \to V$ y $x \to q$, obtenemos $V = \alpha q$. Resulta que existe tal cosa, de hecho es el único de los tres elementos de circuito que podemos entender realmente, porque hemos estudiado un par de placas paralelas y encontramos que si hubiera una carga de cierta cantidad igual y contraria en cada placa, el campo eléctrico entre ellas sería proporcional al tamaño de la carga. De manera que el trabajo realizado al mover una carga unitaria a través de la separación de una placa a la otra es precisamente proporcional a la carga. Este trabajo es la *definición* de diferencia de potencial y es la integral curvilínea del campo eléctrico de una placa a la otra. Resulta que por razones históricas, la constante de proporcionalidad no se llama C sino $1/C$. Pudo haber sido llamada C, pero no lo fue. Tenemos entonces

$$V = q/C. \tag{25.13}$$

La unidad de capacitancia C es el faradio, una carga de un coulomb en cada placa de un capacitor de un faradio produce una diferencia de potencial de un voltio.

Ahí están nuestras analogías y la ecuación correspondiente al circuito oscilante se transforma en lo siguiente, mediante sustitución directa de m por L, x por q, etcétera:

$$m(d^2x/dt^2) + \gamma m(dx/dt) + kx = F, \tag{25.14}$$

$$L(d^2q/dt^2) + R(dq/dt) + q/C = V. \tag{25.15}$$

Ahora bien, todo lo que aprendimos acerca de (25.14) se puede transformar para aplicarlo a (25.15). Cualquier *consecuencia* es la misma, es tan igual, que hay algo brillante que podemos hacer.

Supongan que tenemos un sistema mecánico bastante complicado, no sólo una masa en un resorte, sino que varias masas en varios resortes todos enganchados entre sí. ¿Qué hacemos? ¿Resolverlo? A lo mejor, pero miren, podemos hacer un circuito *eléctrico* ¡que va a tener las mismas ecuaciones que el objeto que estamos tratando de analizar! Por ejemplo, si queríamos analizar una masa en un resorte, ¿por qué no construir un circuito eléctrico la cual empleamos una inductancia proporcional a la masa, una resistencia proporcional al $m\gamma$ correspondiente, $1/C$ que sea proporcional a k, todas en la misma proporción? Entonces este circuito eléctrico va a ser, por supuesto, la analogía exacta del mecánico nuestro en el sentido que cualquier cosa que haga q en respuesta a V (V también se hace corresponder a las fuerzas que están actuando) ¡así lo hará x en respuesta a la fuerza! De manera que si tenemos algo complicado con una gran cantidad de elementos conectados, podemos conectar una cantidad de resistencias, inductancias y capacitancias para *imitar* el sistema mecánico complicado. ¿Cuál es la ventaja de esto? Un problema es tan difícil, o tan fácil como el otro, porque son exactamente equivalentes. La ventaja no es que sea más fácil resolver las *ecuaciones matemáticas* después que descubrimos que tenemos un circuito eléctrico (¡aunque ése *es* el método usado por los ingenieros

eléctricos!), sino que en vez de ello, la verdadera razón para mirar el análogo es que es más fácil *hacer* el circuito eléctrico y *cambiar* algo en el sistema.

Supongan que hemos diseñado un automóvil y queremos saber cuánto se va a sacudir cuando pase por un cierto tipo de camino lleno de baches. Construimos un circuito eléctrico con inductancias que representen la inercia de las ruedas, las constantes de resorte como condensadores para representar los resortes de las ruedas y resistencias para representar los amortiguadores, etc., para las otras partes del automóvil. A continuación necesitamos un camino lleno de baches. De acuerdo, aplicamos un *voltaje* de un generador que represente tal y tal tipo de bache y después miramos cómo se sacude la rueda izquierda midiendo la carga en algún condensador. Habiéndola medido (lo que es fácil de hacer) encontramos que se está sacudiendo mucho. ¿Necesitamos más amortiguamiento o menos? Con una cosa tan complicada como un automóvil, ¿realmente cambiamos el amortiguador y lo resolvemos todo de nuevo? ¡No!, simplemente movemos un dial, el dial número 10 es el amortiguador número tres. Así que introducimos más amortiguamiento. Las sacudidas son peores —de acuerdo, probemos con menos—. Las sacudidas son peores aún, cambiamos la dureza del resorte (dial 17) y ajustamos todas estas cosas *eléctricamente*, simplemente girando una perilla.

Esto se llama *computadora analógica*. Es un dispositivo que imita el problema que queremos resolver, creando otro problema que tiene la misma ecuación, pero en otras circunstancias de la naturaleza y ¡que es mucho más fácil de construir, de medir, de ajustar y de destruir!

25-5 Impedancias en serie y en paralelo

Finalmente, hay un detalle importante que no tiene casi la naturaleza de repaso. Tiene que ver con un circuito eléctrico en el cual hay más de un elemento de circuito. Por ejemplo, cuando tenemos un inductor, una resistencia y un condensador conectados como en la figura 24-2, notamos que toda la carga pasaba por cada uno de los tres, de manera que en un objeto tal conectado tan singularmente, la corriente es la misma en todos los puntos del alambre. Como la corriente es la misma en cada uno, el voltaje a través de R es IR, el voltaje a través de L es $L(dI/dt)$, etc. De manera que la caída total del potencial es la suma de éstos y ello conduce a la ecuación (25.15). Usando números complejos, encontramos que podíamos resolver la ecuación para el movimiento estacionario en respuesta a una fuerza sinusoidal. Encontramos aquí que $\hat{V} = \hat{Z}\hat{I}$. Ahora $\hat{Z}$ se llama la *impedancia* de ese circuito particular. Nos dice que si aplicamos un voltaje sinusoidal $\hat{V}$, obtenemos una corriente $\hat{I}$.

Supongan ahora que tenemos un circuito más complicado que tiene dos partes que por sí mismas tienen una cierta impedancia $\hat{Z}_1$ y $\hat{Z}_2$ y las ponemos en *serie* (figura 25-6a) y aplicamos un voltaje. ¿Qué pasa? Ahora es un poco más complicado, pero si $\hat{I}$ es la corriente a través de $\hat{Z}_1$, la diferencia de potencial a través de $\hat{Z}_1$ es $\hat{V}_1 = \hat{I}\hat{Z}_1$; igualmente, el voltaje a través de $\hat{Z}_2$ es $\hat{V}_2 = \hat{I}\hat{Z}_2$. *La misma corriente pasa por ambos.* Luego el voltaje total es la suma de los voltajes a través de las dos secciones y es igual a $\hat{V} = \hat{V}_1 + \hat{V}_2 = (\hat{Z}_1 + \hat{Z}_2)\hat{I}$. Esto significa que el voltaje en el circuito completo se puede escribir $\hat{V} = \hat{I}\hat{Z}_s$, donde la $\hat{Z}_s$ del sistema combinado en serie es la suma de los dos $\hat{Z}$ de cada parte:

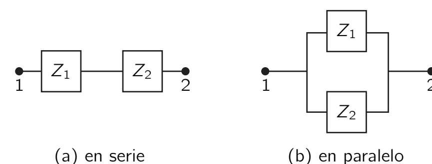


Figura 25-6. Dos impedancias conectadas en serie y en paralelo.

$$\hat{Z}_s = \hat{Z}_1 + \hat{Z}_2. \tag{25.16}$$

Ésta no es la única manera de conectar las cosas. También las podemos conectar de otra manera, llamada conexión en *paralelo* (figura 25-6b). Ahora vemos que un voltaje dado aplicado a los terminales, si los alambres de conexión son conductores perfectos, está aplicado efectivamente a ambas impedancias y va a dar origen a corrientes en cada una independientemente. Luego, la corriente en $\hat{Z}_1$ es igual a $\hat{I}_1 = \hat{V}/\hat{Z}_1$. La corriente en $\hat{Z}_2$ es $\hat{I}_2 = \hat{V}/\hat{Z}_2$. Es el *mismo voltaje*. La corriente total que se entrega a los terminales es la *suma* de las corrientes en

las dos secciones, $\hat{I} = \hat{V}/\hat{Z}_1 + \hat{V}/\hat{Z}_2$. Esto se puede escribir:

$$\hat{V} = \frac{\hat{I}}{(1/\hat{Z}_1) + (1/\hat{Z}_2)} = \hat{I}\hat{Z}_p.$$

Así

$$1/\hat{Z}_p = 1/\hat{Z}_1 + 1/\hat{Z}_2. \tag{25.17}$$

Circuitos más complicados pueden simplificarse a veces tomando partes de ellos, trabajando con las impedancias sucesivas de las partes y combinando el circuito paso a paso, usando las reglas anteriores. Si tenemos cualquier tipo de circuito con muchas impedancias conectadas de todas las maneras posibles y si incluimos los voltajes en forma de pequeños generadores sin impedancia (cuando pasamos carga a través de él, el generador agrega un voltaje V), entonces se aplican los siguientes principios: (1) En cualquier nudo la suma de las corrientes hacia el nudo es cero. O sea, toda la corriente que entra debe salir. (2) Si llevamos una carga alrededor de cualquier malla, volviendo al punto de partida, el trabajo neto realizado es cero. Estas reglas se llaman *leyes de Kirchhoff* para circuitos eléctricos. Su aplicación sistemática a circuitos complicados a menudo simplifica su análisis. Los mencionamos aquí junto con las ecuaciones (25.16) y (25.17), en el caso de que ustedes ya hayan encontrado tales circuitos que necesitan analizar en el trabajo de laboratorio. Van a ser discutidas de nuevo con mayor detalle el próximo año.

26

Óptica: el principio del tiempo mínimo

26-1 La luz

Éste es el primero de una serie de capítulos sobre el tema de la *radiación electromagnética.* La luz, con la cual vemos, es sólo una pequeña parte de un vasto espectro del mismo tipo de cosas, distinguiéndose las diversas partes de este espectro por valores diferentes de una cierta cantidad que varía. Esta cantidad variable podría llamarse la “longitud de onda”. A medida que ésta varía en el espectro visible, la luz aparentemente cambia de color de rojo a violeta. Si exploramos el espectro sistemáticamente, desde longitudes de ondas largas hacia las más cortas, empezaríamos con lo que se llama comúnmente *ondas de radio.* Las ondas de radio son asequibles técnicamente en un gran intervalo de longitudes de onda, algunas aún más largas que las usadas en las radiodifusiones regulares; las difusiones radiales regulares tienen longitudes de ondas correspondientes a unos 500 metros. A continuación están las así llamadas “ondas cortas”, por ejemplo, ondas de radar, ondas milimétricas, etc. No existe ningún límite definido entre uno y otro intervalo de longitudes de ondas, porque la naturaleza no se nos presenta con límites bien definidos. El número asociado con un nombre dado para las ondas es aproximado solamente y también lo son, por supuesto, los nombres que damos a los diferentes intervalos.

Después, de un largo camino por las ondas milimétricas, llegamos a la región que llamamos *infrarroja* y, por lo tanto, al espectro visible. Avanzando hacia el otro lado, llegamos a la región que se llama *ultravioleta.* Donde termina el ultravioleta empiezan los rayos X, pero no podemos definir precisamente dónde sucede esto; es aproximadamente a 10^{-8} m, o $10^{-2}\mu$. Éstos son rayos X “blandos”, luego están los rayos ordinarios y los rayos X muy duros, luego los rayos γ, etc., para valores cada vez menores de esta dimensión llamada longitud de onda.

Dentro de esta vasta gama de longitudes de onda, hay tres o más regiones de aproximación que son especialmente interesantes. En una de éstas, existe una condición en la cual las longitudes de onda que intervienen son muy pequeñas comparadas con las dimensiones del equipo disponible para su estudio; además, las energías de los fotones, usando la teoría cuántica, son pequeñas comparadas con la sensibilidad energética del equipo. En estas condiciones podemos hacer una primera aproximación por un método llamado *óptica geométrica.* Si, por otro lado, las longitudes de ondas son comparables a las dimensiones del equipo, lo que es difícil de obtener con la luz visible pero más fácil con ondas de radio, y si las energías de los fotones son todavía despreciables, pues son chicas, entonces se puede hacer una aproximación muy útil estudiando el comportamiento de las ondas, aun sin tomar en cuenta la mecánica cuántica. Este método se basa en la *teoría clásica de la radiación electromagnética*, que va a ser discutida en un capítulo más adelante. A continuación, si vamos a las longitudes de ondas más cortas, donde podemos despreciar el carácter ondulatorio, pero donde los fotones tienen una energía muy *grande* comparada con la sensibilidad del equipo, las cosas se hacen sencillas de nuevo. Ésta es la imagen simple del *fotón* que vamos a describir sólo muy aproximadamente. La imagen completa, que unifica todo en un solo modelo, no nos será asequible por mucho tiempo.

En este capítulo nuestra discusión se limita a la región de la óptica geométrica, en la cual nos olvidamos del carácter ondulatorio y fotónico de la luz, lo que va a ser explicado a su debido tiempo. Ni siquiera nos preocupamos de decir lo que *es* la luz, sino sólo de averiguar *cómo se comporta* en una escala grande comparada con las dimensiones de interés. Todo esto debe ser dicho para poner énfasis en

que lo que vamos a decir es una mera aproximación; éste es uno de los capítulos que vamos a tener que "desaprender" de nuevo. Pero lo vamos a desaprender muy pronto, porque vamos a seguir casi inmediatamente con un método más preciso.

Aunque la óptica geométrica es sólo una aproximación, es de una gran importancia técnica y de gran interés histórico. Vamos a presentar este tema más históricamente que otros para dar una idea del desarrollo de una teoría física o de una idea física.

Primero, la luz es, por supuesto, familiar para todo el mundo y ha sido familiar desde tiempo inmemorial. Ahora bien, un problema es: ¿mediante qué proceso *vemos* la luz? Ha habido muchas teorías, pero finalmente se asentó una según la cual es que hay algo que entra al ojo —que rebota de los objetos y entra al ojo—. Hemos escuchado esa idea por tanto tiempo que la aceptamos y es casi imposible para nosotros darnos cuenta de que hombres muy inteligentes hayan propuesto teorías contrarias —algo sale del ojo y palpa el objeto, por ejemplo—. Algunas otras observaciones importantes son que cuando la luz va de un lugar a otro, va en *líneas rectas*, si no hay nada en el camino, y que los rayos no parecen interferir unos con otros. Esto es, la luz se entrecruza en todas direcciones en la pieza, pero la luz que está pasando por nuestra línea de visión no afecta la luz que nos llega de algún objeto. Esto fue una vez un argumento poderosísimo en contra de la teoría corpuscular, fue usado por Huygens. Si la luz fuera como muchas flechas disparadas, ¿cómo podrían otras flechas pasar entre ellas tan fácilmente? Tales argumentos filosóficos no son de mucho peso. ¡Uno siempre podría decir que la luz está formada por flechas que pasan a través de ellas mismas!

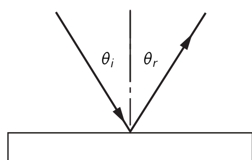


Figura 26-1. El ángulo de incidencia es igual al ángulo de reflexión.

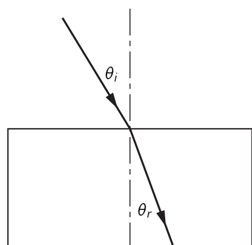


Figura 26-2. Un rayo de luz se refracta cuando pasa de un medio a otro.

Tabla 26-1

Ángulo en el aire	Ángulo en el agua
10°	8°
20°	15-1/2°
30°	22-1/2°
40°	29°
50°	35°
60°	40-1/2°
70°	45-1/2°
80°	50°

26-2 Reflexión y refracción

La discusión anterior es suficiente sobre el *concepto* básico de la óptica geométrica —ahora debemos ir un poco más allá en los aspectos cuantitativos—. Hasta ahora tenemos a la luz yendo en línea recta entre dos puntos solamente; estudiemos ahora el comportamiento de la luz cuando golpea diversos materiales. El objeto más simple es un espejo y la ley para el espejo es que cuando la luz golpea el espejo, no sigue en línea recta, sino que rebota del espejo en una nueva línea recta que cambia cuando cambiamos la inclinación del espejo. La pregunta para los antiguos era: ¿cuál es la relación entre los dos ángulos que entran en juego? Ésta es una relación muy simple, descubierta hace mucho tiempo. La luz que choca con un espejo viaja de tal manera que los dos ángulos entre cada rayo y el espejo son iguales. Por alguna razón se acostumbra medir los ángulos desde la normal a la superficie del espejo. De manera que la llamada ley de reflexión es:

$$\theta_i = \theta_r. \tag{26.1}$$

Ésta es una proposición bastante simple, pero se encuentra un problema más difícil cuando la luz va desde un medio a otro, por ejemplo, desde el aire al agua; aquí también vemos que no va en línea recta. En el agua el rayo tiene una inclinación respecto a su trayectoria en el aire; si cambiamos el ángulo θ_i de manera que se acerque a la vertical, entonces el ángulo "en que se quiebra" no es tan grande. Pero si inclinamos el rayo de luz en un ángulo considerable, entonces la desviación es muy grande. La pregunta es: ¿cuál es la relación entre uno y otro ángulo? Esto intrigó también a los antiguos por mucho tiempo ¡y aquí nunca encontraron la respuesta! Es, sin embargo, uno de los pocos lugares de toda la física griega donde uno puede encontrar algunos resultados experimentales registrados. Claudio Ptolomeo hizo una lista del ángulo en el agua para cada uno de un número de ángulos diferentes en el aire. La tabla 26-1 muestra los ángulos en el aire en grados y el ángulo correspondiente medido en el agua. (Comúnmente se dice que los sabios griegos no hacían nunca experimentos. Pero sería imposible obtener esta tabla de valores sin conocer la ley duradera, excepto mediante el experimento. Debe anotarse, sin embargo, que éstos no representan

medidas cuidadosas independientes para cada ángulo, sino sólo algunos números interpolados de unas pocas medidas, porque todos ellos se ajustan perfectamente a una parábola.)

Éste es entonces uno de los pasos más importantes en el desarrollo de una ley física: primero observamos un efecto, luego lo medimos y lo registramos en una tabla, luego tratamos de encontrar la *regla* mediante la cual una cosa está conectada con la otra. La tabla numérica anterior fue hecha en el año 140 d.C., ¡pero no fue sino en 1621 cuando alguien encontró finalmente la regla que conecta a los dos ángulos! La regla, encontrada por Willebrord Snell, un matemático holandés, es como sigue: si θ_i es el ángulo en el aire y θ_r el ángulo en el agua, entonces resulta que el seno de θ_i es igual a un cierto múltiplo constante del seno de θ_r:

$$\operatorname{sen}\theta_i = n \operatorname{sen}\theta_r. \tag{26.2}$$

Para el agua el número n es aproximadamente 1.33. La ecuación (26.2) se llama *ley de Snell*; nos permite *predecir* cómo se va a inclinar cuando va desde el aire al agua. La tabla 26-2 muestra los ángulos en el aire y en el agua de acuerdo con la ley de Snell. Nótese la notable similitud con la lista de Ptolomeo.

Tabla 26-2

Ángulo en el aire	Ángulo en el agua
10°	7-1/2°
20°	15°
30°	22°
40°	29°
50°	35°
60°	40-1/2°
70°	45°
80°	48°

26-3 El principio de Fermat del tiempo mínimo

Ahora bien, en un mayor desarrollo de la ciencia queremos más que una fórmula. Primero tenemos una observación, luego números que medimos, luego tenemos una ley que resume todos los números. Pero la verdadera *gloria* de la ciencia es que *podemos encontrar una manera de pensar* tal que la ley sea *evidente*.

La primera manera de pensar que hizo evidente la ley del comportamiento de la luz fue descubierto por Fermat cerca de 1650, y se llama *principio del tiempo mínimo* o *principio de Fermat*. Su idea es ésta: que de todos los posibles caminos que puede tomar para ir de un punto a otro, la luz toma el camino que requiere el *tiempo más corto*.

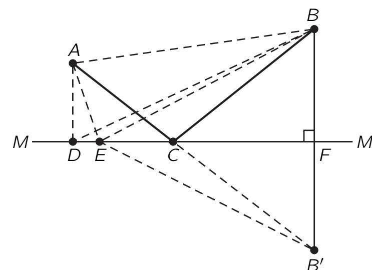


Figura 26-3. Ilustración del principio del tiempo mínimo.

Mostremos primero que esto es cierto para el caso del espejo, que este principio contiene tanto la ley de la propagación rectilínea como la ley del espejo. ¡De manera que nuestra comprensión está aumentando! Tratemos de encontrar la solución para el siguiente problema. En la figura 26-3 se muestran dos puntos A y B y un espejo plano MM'. ¿Cuál es la manera de ir de A a B en el tiempo más corto? ¡La respuesta es ir directamente de A a B! Pero si agregamos una regla extra que la luz tiene que *chocar con el espejo* y volver en el tiempo más corto, la respuesta no es tan sencilla. Una manera podría ser ir lo más rápidamente posible al espejo y después ir a B por el camino ADB. Por supuesto, obtenemos entonces un camino largo DB. Si nos movemos un poco hacia la derecha, a E, aumentamos ligeramente la primera distancia, pero *disminuimos* mucho la segunda y así el largo del camino total, y por lo tanto el tiempo de viaje, es menor. ¿Cómo se puede encontrar el punto C para el cual el tiempo es el más corto? Lo podemos encontrar de una manera muy bonita mediante un truco geométrico.

Construimos al otro lado de MM' un punto artificial B' que está a la misma distancia debajo del plano MM' que el punto B está sobre el plano. Entonces dibujamos la línea EB'. Ahora bien, como BFM es un ángulo recto y $BF = FB'$, EB es igual a EB'. Luego la suma de las dos distancias $AE + EB$, que es proporcional al tiempo que se va a demorar si la luz viaja con velocidad constante, también es la suma de las dos longitudes $AE + EB'$. Por lo tanto el problema se convierte en: ¿cuándo es mínima la suma de estas dos longitudes? La respuesta es fácil: ¡cuando la línea pasa por el punto C como una *línea recta* de A a B'! En otras palabras, tenemos que encontrar el punto a partir del cual vamos hacia el punto artificial y ése será el correcto. Ahora, si ACB' es una línea recta, entonces el ángulo BCF es igual al ángulo $B'CF$ y por lo tanto al ángulo ACM. Por lo tanto, la afirmación de que el ángulo de incidencia es igual al ángulo de reflexión

es equivalente a la afirmación de que la luz va al espejo de manera tal que vuelve al punto B en el *menor tiempo posible*. Originalmente Herón de Alejandría afirmó que la luz viaja de tal manera que va al espejo y al otro punto en la menor *distancia* posible, así que no es una teoría moderna. Fue esto lo que inspiró a Fermat a sugerir que tal vez la refracción funcionaba con bases similares. Pero para la refracción, la luz evidentemente no usa el camino de menor *distancia*; así, Fermat ensayó la idea de que toma el *tiempo* más corto.

Antes que prosigamos analizando la refracción, debiéramos hacer una observación más acerca del espejo. Si tenemos una fuente de luz en el punto B y envía luz hacia el espejo, vemos que la luz que va hacia A desde el punto B llega a A exactamente de la misma manera como habría llegado a A si hubiera habido un objeto en B' y *ningún* espejo. Por supuesto que el ojo detecta solamente la luz que entra físicamente, en particular, si tenemos un objeto en B y un espejo que hace que la luz llegue al ojo exactamente de la misma manera como habría llegado al ojo si el objeto hubiera estado en B', entonces el sistema ojo-cerebro interpreta eso, suponiendo que no sabe mucho, como si *hubiera* un objeto en B'. De manera que la ilusión de que hay un objeto detrás del espejo se debe simplemente a que la luz que entra al ojo está entrando de la misma manera, físicamente, como habría entrado si hubiera *habido* un objeto detrás (con excepción de la suciedad en el espejo y de nuestro conocimiento de la existencia del espejo, etc., lo que se corrige en el cerebro).

Demostremos ahora que el principio del tiempo mínimo va a dar la leyes de Snell de la refracción. Debemos, sin embargo, hacer una suposición acerca de la velocidad de la luz en el agua. Vamos a suponer que la velocidad de la luz en el agua es menor que la velocidad de la luz en el aire, en un cierto factor n.

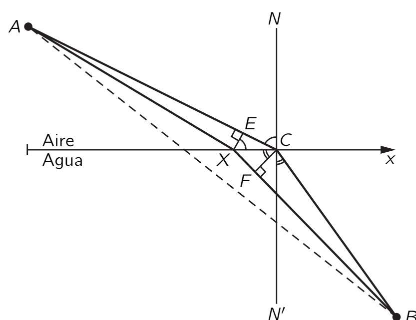


Figura 26-4. Ilustración del principio de Fermat para la refracción.

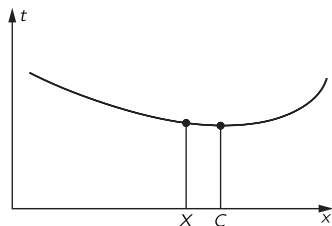


Figura 26-5. El tiempo mínimo corresponde al punto C, pero puntos cercanos corresponden a casi el mismo tiempo.

En la figura 26-4 nuestro problema es de nuevo ir de A a B en *el tiempo más corto*. Para demostrar que lo mejor no es ir simplemente en línea recta, supongamos que una hermosa jovencita se ha caído de un bote y está gritando pidiendo ayuda en el agua en el punto B. La línea marcada X es la orilla. Nosotros estamos en un punto A en tierra y vemos el accidente y podemos correr y también nadar. Pero podemos correr más rápido de lo que podemos nadar. ¿Qué hacemos? ¿Vamos en línea recta? (¡Sí, sin duda!) Sin embargo, si usáramos un poco más de inteligencia nos daríamos cuenta de que es ventajoso seguir una distancia un poco mayor por tierra para disminuir la distancia en el agua, porque nos movemos más lentamente en el agua. (¡Siguiendo esta línea de razonamiento, diríamos que lo correcto sería *calcular* muy cuidadosamente lo que debe hacerse!) En fin, tratemos de demostrar que la solución final del problema es el camino ACB, y que este camino toma el menor de los tiempos posibles. Si es el camino más corto, eso significa que si tomamos cualquier otro va a ser más largo. De manera que si graficáramos el tiempo que demora en función de la posición del punto X, obtendríamos una curva algo parecida a la que se muestra en la figura 26-5 donde el punto C corresponde al menor de los tiempos posibles. Esto significa que si movemos el punto X hasta puntos *cercanos* a C, en primera aproximación no hay esencialmente *ningún cambio* en el tiempo, porque la inclinación es cero en el fondo de la curva. Así, pues, nuestra manera de encontrar la ley va a ser considerar que movemos el lugar en una pequeña cantidad y exigir que no haya esencialmente ningún cambio en el tiempo. (Por supuesto que hay un cambio infinitesimal de *segundo* orden; deberíamos tener un aumento positivo para desplazamientos en cualquier dirección desde C.) De manera que consideramos un punto vecino X y calculamos cuánto tomaría en ir de A a B por los dos caminos y comparamos el nuevo camino con el antiguo. Esto es muy fácil de hacer. Queremos que la diferencia sea, por supuesto, casi cero si la distancia XC es pequeña. Primero, fíjense en el camino en tierra. Si dibujamos una perpendicular XE vemos que este camino está acortado en una cantidad EC. Digamos que ganamos por no tener que ir esa distancia adicional. Por otro lado, en el agua, dibujando una perpendicular correspondiente CF, encontramos que tenemos que avanzar la distancia adicional XF, y eso es lo que perdemos. O sea que en *tiempo*, ganamos el tiempo que habría tomado avanzar la distancia EC, pero perdemos el tiempo que habría tomado avanzar la distancia

XF. Estos tiempos deben ser iguales, ya que en primera aproximación no debe haber cambio en el tiempo. Pero suponiendo que en el agua la velocidad es $1/n$ veces la del aire, debemos tener:

$$EC = n \cdot XF. \tag{26.3}$$

Por lo tanto, vemos que cuando tenemos el punto preciso, $XC \operatorname{sen} EXC = n \cdot XC \operatorname{sen} XCF$ o, simplificando el largo de la hipotenusa común XC y notando que:

$$EXC = ECN = \theta_i \quad \text{y} \quad XCF \approx BCN' = \theta_r \quad (\text{cuando } X \text{ está cerca de } C),$$

tenemos

$$\operatorname{sen}\theta_i = n \operatorname{sen}\theta_r. \tag{26.4}$$

Por consiguiente, para ir de un punto a otro en el tiempo mínimo cuando el cociente entre las velocidades es n, la luz debe entrar en un ángulo cuyo cociente entre los senos de los ángulos θ_i y θ_r sea el cociente entre las velocidades en los dos medios.

26-4 Aplicaciones del principio de Fermat

Consideremos ahora alguna de las consecuencias interesantes del principio del tiempo mínimo. Primero está el principio de reciprocidad. Si al ir de A a B hemos encontrado el camino de tiempo mínimo, al ir en la dirección opuesta (suponiendo que la luz va con la misma velocidad en cualquier dirección) el tiempo más corto corresponderá al mismo camino y, por lo tanto, si la luz puede ser enviada en un sentido, puede ser enviada en el otro.

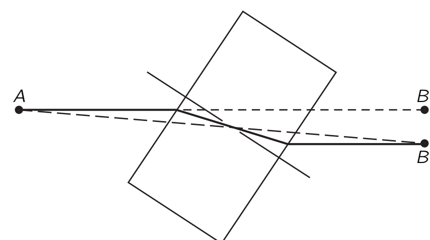


Figura 26-6. Un rayo de luz se desvía cuando pasa por un bloque transparente.

Un ejemplo interesante es un bloque de vidrio con caras paralelas planas formando un ángulo con el haz de luz. La luz, al ir a través del bloque desde el punto A al punto B (figura 26-6), no va en línea recta, sino que disminuye el tiempo en el bloque haciendo el ángulo en el mismo menos inclinado, aunque pierde un poco en el aire. El rayo se desplaza sencillamente paralelo a sí mismo porque los ángulos de entrada y de salida son los mismos.

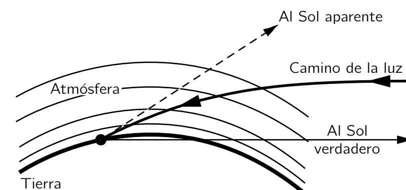


Figura 26-7. Cerca del horizonte el Sol aparente está más alto que el Sol verdadero, cerca de 1/2 grado.

Un tercer fenómeno interesante es que cuando vemos ponerse el Sol, ¡ya está bajo el horizonte! No *parece* como si estuviera debajo del horizonte, pero está (figura 26-7). La atmósfera de la Tierra es enrarecida arriba y densa abajo. La luz viaja más lentamente en el aire que en el vacío, y así la luz del Sol puede llegar al punto S más allá del horizonte más rápidamente si en vez de ir simplemente en línea recta, evita las regiones densas donde va más lentamente, atravesándolas con una mayor inclinación. Cuando parece ponerse tras el horizonte, realmente está hace rato bajo el horizonte. Otro ejemplo de este fenómeno es el espejismo que uno ve a menudo cuando maneja en caminos calurosos. Uno ve "agua" en el camino, pero cuando llega ahí ¡está tan seco como el desierto! Lo que estamos viendo realmente es la luz del cielo "reflejada" en el camino, la luz del cielo, con rumbo al camino, puede terminar en el ojo como se muestra en la figura 26-8. ¿Por qué? El aire está muy caliente justo sobre el camino, pero está más frío más arriba. El aire caliente está más expandido que el aire frío y es más enrarecido y es menor la disminución de la velocidad de la luz. O sea, la luz va más rápido en la región caliente que en la región fría. Por lo tanto, la luz en vez de decidir irse de una manera directa, tiene también un camino de tiempo mínimo, por el cual entra por un rato en la región donde se mueve más rápido a fin de ahorrar tiempo. Así, pues, puede seguir una curva.



Figura 26-8. Un espejismo

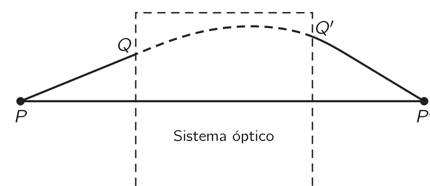


Figura 26-9. Una "caja negra" óptica.

Como otro ejemplo importante del principio de tiempo mínimo, supongan que quisiéramos disponer una situación en la cual tenemos que toda la luz que sale de un punto P se recolecta en otro punto P' (figura 26-9). Eso significa, por supuesto, que la luz puede ir en una línea recta de P a P'. Eso está bien, pero ¿cómo podemos arreglarnos para que no solamente vaya derecho, sino que la luz que parte de P a Q también termine en P'? Queremos juntar toda la luz en lo que llamamos *foco*. ¿Cómo? Si la luz toma siempre el camino de tiempo

mínimo, entonces ciertamente no debería querer seguir todos esos caminos. ¡La única manera para que la luz pueda estar perfectamente satisfecha de tomar varios caminos adyacentes es hacer aquellos tiempos *exactamente iguales*! De otra manera, seleccionaría el de tiempo mínimo. Por lo tanto, el problema de hacer un sistema de enfoque es sencillamente arreglar un dispositivo tal ¡que le toma a la luz el mismo tiempo para ir por *todos* los caminos diferentes!

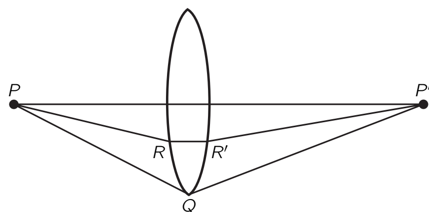


Figura 26-10. Un sistema óptico de enfoque.

Esto es fácil de hacer. Supongan que tuviéramos un pedazo de vidrio en el cual la luz va más despacio que en el aire (figura 26-10). Ahora consideren un rayo que va por el aire por el camino PQP'. Ese es un paso más largo que desde P directamente a P' y sin duda demora un tiempo mayor. Pero si introdujéramos un pedazo de vidrio del grosor justo (después averiguaremos qué grosor) ¡podría compensar exactamente el exceso de tiempo que demoraría la luz yendo oblicua! En este caso podemos hacer que el tiempo que la luz demora en atravesar derecho sea el mismo que el tiempo que demora en ir por el camino PQP'. Igualmente, si tomamos un rayo $PRR'P'$ que está algo inclinado, no es tan largo como PQP' y no tenemos que compensar tanto como para el recto, pero tenemos que compensar algo. Terminamos con un pedazo de vidrio como el de la figura 26-10. Con esta forma toda la luz que viene de P va a P'. Esto, por supuesto, es bien conocido por nosotros y llamamos a tal dispositivo *lente* convergente. En el capítulo siguiente vamos a calcular realmente qué forma debe tener la lente para hacer un enfoque perfecto.

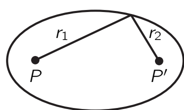


Figura 26-11. Un espejo elipsoidal.

Consideren otro ejemplo, supongan que deseamos disponer algunos espejos de manera que la luz de P vaya siempre a P' (figura 26-11). Por cualquier camino ella va a algún espejo y vuelve, y todos los tiempos deben ser iguales. Aquí siempre la luz viaja en el aire de manera que el tiempo y la distancia son proporcionales. Por lo tanto, la afirmación que todos los tiempos son iguales es igual a la afirmación que la distancia total es la misma. Por lo tanto, la suma de las dos distancias r_1 y r_2 debe ser una constante. Una *elipse* es la curva que tiene la propiedad de que la suma de las distancias desde dos puntos es constante para todo punto sobre la elipse; así, podemos estar seguros de que la luz de un foco va a ir al otro.

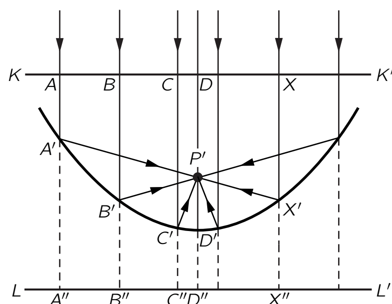


Figura 26-12. Un espejo en forma de parábola.

El mismo principio funciona para concentrar la luz de una estrella. El gran telescopio de Palomar de cinco metros está construido según el siguiente principio. Imaginen una estrella alejada miles de millones de kilómetros, desearíamos lograr que toda la luz que entra llegue a un foco. Por cierto que no podemos dibujar los rayos que llegan hasta la estrella; ahora bien, de todos modos deseamos verificar si los tiempos son iguales. Por supuesto, sabemos que cuando los diversos rayos han llegado a un plano KK', perpendicular a los rayos, todos los tiempos en este plano son iguales (figura 26-12). Los rayos deben entonces bajar hasta el espejo y proseguir a P' en tiempos iguales. Esto es, debemos encontrar una curva que tenga la propiedad de que la suma de las distancias $XX' + X'P'$ sea una constante, cualquiera que sea la elección X. Una manera fácil de encontrarla es prolongar la línea XX' hasta un plano LL'. Ahora bien, si disponemos nuestra curva de manera que $A'A'' = A'P'$, $B'B'' = B'P'$, $C'C'' = C'P'$, etc., tendremos nuestra curva, porque entonces, por supuesto, $AA' + A'P' = AA' + A'A''$ va a ser constante. Así, pues, nuestra curva es el lugar geométrico de todos los puntos equidistantes de una línea y un punto. Esa curva se llama *parábola*, el espejo se hace en forma de parábola.

Los ejemplos anteriores ilustran el principio con el cual pueden ser diseñados esos dispositivos ópticos. Las curvas exactas pueden ser calculadas usando el principio según el cual, para enfocar perfectamente, los tiempos de viaje deben ser exactamente iguales para todos los rayos de luz, al mismo tiempo que son menores que para cualquier otro camino vecino.

Vamos a discutir más estos dispositivos ópticos de enfoque en el capítulo siguiente; discutamos ahora el desarrollo subsiguiente de la teoría. Cuando se desarrolla un nuevo principio teórico, como el principio del tiempo mínimo, nuestra primera inclinación sería querer decir: “Bueno, eso es muy bonito, es delicioso, pero la pregunta es: ¿ayuda en algo a comprender la física?” Alguien podría decir: “Sí, ¡vean cuántas cosas podemos comprender ahora!” Otro dice:

"Muy bien, pero yo puedo comprender los espejos también. Necesito una curva tal que todo plano tangente forme ángulos iguales con los dos rayos. Puedo idear una lente también, porque cada rayo que llega a él se dobla en ángulo por la ley de Snell". Evidentemente la afirmación de que el tiempo es mínimo y la afirmación de que los ángulos son iguales en la reflexión y que los senos de los ángulos son proporcionales en la refracción son lo mismo. De manera que, ¿es sólo una pregunta filosófica o estética? Puede haber argumentos para ambos bandos.

Sin embargo, la importancia de un principio poderoso *es que predice nuevas cosas.*

Es fácil demostrar que hay una cantidad de cosas nuevas predichas por el principios de Fermat. Primero, supongamos que hay *tres* medios, vidrio, agua y aire; realizamos un experimento de refracción y medimos el índice n de un medio respecto al otro. Llamemos n_{12} el índice del aire (1) respecto al agua (2), n_{13} el índice del aire (1) respecto al vidrio (3). Si midiéramos agua respecto a vidrio, deberíamos encontrar otro índice que llamaremos n_{23}. Pero no hay ninguna razón *a priori* porque habría de haber alguna conexión entre n_{12}, n_{13} y n_{23}. Por otro lado, de acuerdo con la idea del tiempo mínimo, existe una relación bien definida. El índice n_{12} es el cociente entre dos cosas, la velocidad en el aire y la velocidad en el agua; n_{13} es el cociente entre la velocidad en el aire y la velocidad en el vidrio, n_{23} es el cociente entre la velocidad en el agua y la velocidad en el vidrio. Luego simplificamos el aire y obtenemos:

$$n_{23} = \frac{v_2}{v_3} = \frac{v_1/v_3}{v_1/v_2} = \frac{n_{13}}{n_{12}}. \tag{26.5}$$

En otras palabras, *predecimos* que el índice para un nuevo par de materiales se puede obtener de los índices de los materiales individuales, ambos respecto al aire o respecto al vacío. Así que, si medimos la rapidez de la luz en todos los materiales y de ahí obtenemos un sólo número para cada material, a saber, su índice relativo al vacío, llamado n_i (n_i es la velocidad en el vacío relativa a la velocidad en el aire, etc.), entonces nuestra fórmula es fácil. El índice para dos materiales cualesquiera i y j es:

$$n_{ij} = \frac{v_i}{v_j} = \frac{n_j}{n_i}. \tag{26.6}$$

Usando sólo la ley de Snell, no hay base para una predicción de este tipo.* Pero, por supuesto, esta predicción funciona. La relación (26-5) se conoció desde muy temprano y fue un poderoso argumento a favor del principio del tiempo mínimo.

Otro argumento para el principio del tiempo mínimo, otra predicción, es que si *medimos* la velocidad de la luz en el agua, va a ser menor que en el aire. Ésta es una predicción de un tipo totalmente diferente. Es una predicción brillante, porque todo lo que hemos medido hasta ahora son *ángulos;* aquí tenemos una predicción teórica, que es bastante diferente de las observaciones de las cuales Fermat dedujo la idea del tiempo mínimo. ¡Resulta, de hecho, que la velocidad en el agua *es* menor que la velocidad en el aire en la proporción justa que se necesita para tener el índice correcto!

26-5 Un enunciado más preciso del principio de Fermat

Realmente debemos hacer el enunciado del principio del tiempo mínimo un poco más preciso. No fue expuesto correctamente más arriba. Se le llama *incorrectamente* principio del tiempo mínimo y hemos seguido con la descripción incorrecta por conveniencia, pero ahora debemos ver cuál es el enunciado correcto. Supongan que tuviéramos un espejo como en la figura 26-3. ¿Qué le hace pensar a la luz que tiene que ir al espejo? El camino del tiempo *mínimo* es claramente AB.

* Aunque se puede deducir si se hace la hipótesis adicional de que al agregar una capa de una sustancia a la superficie de otra no cambia el posible ángulo de refracción en el último material.

Así, algunas personas podrían decir: "A veces es un tiempo máximo". ¡*No es* un tiempo máximo porque ciertamente un camino curvado tomaría un tiempo aún más largo! El enunciado correcto es el siguiente: un rayo que va en cierto camino particular tiene la propiedad de que si hacemos un pequeño cambio (digamos un corrimiento de 1 %) en el rayo de cualquier manera que sea, digamos en el lugar en el cual llega al espejo o en la forma de la curva o cualquier cosa, *no* va a haber cambios de primer orden en el tiempo, va a haber sólo un cambio de *segundo* orden en el tiempo. En otras palabras, el principio es que la luz toma un camino tal que hay muchos otros caminos vecinos que demoran casi exactamente el *mismo* tiempo.

La siguiente es otra dificultad con el principio del tiempo mínimo, es una que las personas que no gustan de este tipo de teorías nunca pudieron aceptar. Con la teoría de Snell podemos "entender" la luz. La luz va caminando, ve una superficie, se curva porque hace algo en la superficie. La idea de causalidad, de que va de un punto a otro y a otro, etc., es fácil de entender. Pero el principio del tiempo mínimo es un principio filosófico completamente diferente acerca de cómo funciona la naturaleza. En vez de decir que es algo causal que cuando hacemos algo, otra cosa sucede, etc., dice esto: nosotros establecemos la situación y la *luz* decide cuál es el tiempo más corto o el extremo, y elige ese camino. Pero ¿*qué* hace, *cómo* lo averigua? ¿*Huele* ella los caminos vecinos y los compara entre sí? La respuesta es sí, lo hace, en cierto modo. Ése es el aspecto que no se conoce, por supuesto, en la óptica geométrica y que está encerrado en el concepto de *longitud de onda*; la longitud de onda nos dice aproximadamente cuán lejos la luz debe "oler" el camino para controlarlo. Es difícil demostrar este hecho en una escala grande con la luz, porque las longitudes de onda son tan terriblemente cortas. Pero con las ondas de radio, digamos ondas de 3 cm, las distancias que están controlando las ondas de radio son más grandes. Si tenemos una fuente de ondas de radio, un detector y una rendija, como en la figura 26-13, los rayos van, por supuesto, de S a D porque es una línea recta, y si cerramos un poco la rendija, está bien —todavía pasan—. Pero si ahora movemos el detector a un lado hacia D', las ondas no van a ir a través de la rendija ancha de S a D' porque verifican varios caminos vecinos y dicen: "No, mi amigo, todos ellos corresponden a tiempos diferentes". Por otro lado, si *impedimos* a la radiación que verifique los caminos cerrando la rendija hasta una fisura muy fina, entonces no hay más que un camino disponible ¡y la radiación lo toma! ¡Con una rendija angosta llega a D' más radiación que la que llega con una rendija ancha!

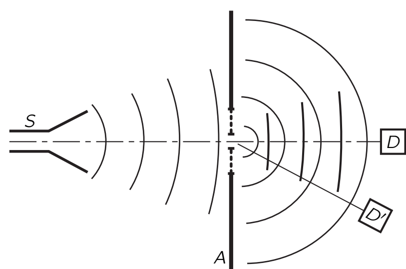


Figura 26-13. El pasaje de una onda de radio a través de una rendija estrecha.

Uno puede hacer lo mismo con la luz, pero es difícil demostrarlo a una escala grande. El efecto se puede ver en las siguientes condiciones simples. Encuentren una luz pequeña y brillante, digamos una lámpara no esmerilada en un farol callejero lejano, o la reflexión del Sol en el parachoques curvado de un automóvil. Entonces pongan dos dedos frente a un ojo de manera de mirar por la hendidura y estrujen la luz hasta cero muy suavemente. Van a ver que la imagen de la luz, que era un pequeño punto antes, se hace bastante alargada y aun se estira en una larga línea. La razón es que los dedos están muy juntos, y la luz que se supone que viene en línea recta se abre en un ángulo, de manera que cuando llega al ojo entra desde varias direcciones. Además notarían, si son muy cuidadosos, máximos laterales, muchas franjas a lo largo de los bordes también. Más aún, todo está coloreado. Todo esto va a ser explicado a su debido tiempo, pero por el momento es una demostración de que la luz no siempre va en líneas rectas, y es una que se realiza muy fácilmente.

26-6 Cómo funciona

Finalmente, damos una visión muy somera de lo que verdaderamente sucede, de cómo el todo funciona realmente a partir de lo que ahora creemos correcto, el punto de vista preciso de la dinámica cuántica, pero por cierto descrito sólo cualitativamente. Al seguir la luz de A a B en la figura 26-3, encontramos que la luz no parece estar en forma de ondas. Por el contrario, los rayos parecen estar hechos de fotones y realmente producirían clics en un contador de fotones, si

es que estuviéramos usando uno. El brillo de la luz es proporcional al número promedio de fotones que llega por segundo, y lo que calculamos es la *probabilidad* de que un fotón llegue de A a B, digamos, al golpear el espejo. La *ley* de esa probabilidad es muy extraña. Tomen cualquier camino y encuentren el tiempo para ese camino, luego formen un número complejo o dibujen un pequeño vector complejo $\rho^{i\theta}$ *cuyo ángulo* θ *sea proporcional al tiempo.* El número de vueltas por segundo es la frecuencia de la luz. Ahora tomen otro camino, tiene, por ejemplo, un tiempo diferente, de manera que su vector está girado en un ángulo diferente —siendo el ángulo siempre proporcional al tiempo—. Tomen *todos* los caminos disponibles y agreguen un pequeño vector por cada uno, ¡entonces la respuesta es que la probabilidad de llegada del fotón es proporcional al cuadrado del largo del vector final, desde el principio hasta el final!

Mostremos ahora cómo esto implica el principio del tiempo mínimo para un espejo. Consideramos todos los rayos, todos los caminos posibles ADB, AEB, ACB, etc., en la figura 26-3. El camino ADB hace una cierta contribución pequeña, pero el camino siguiente AEB demora un tiempo muy distinto, de manera que su ángulo θ es bastante diferente. Digamos que el punto C corresponde al tiempo mínimo, donde si cambiamos los caminos, los tiempos no cambian. Así que por un rato los tiempos cambian y entonces empiezan a variar cada vez menos a medida que nos acercamos al punto C (figura 26-14). De manera que las flechas que tenemos que sumar vienen con casi exactamente el mismo ángulo por un rato cerca de C y entonces gradualmente el tiempo empieza a aumentar de nuevo y las fases giran hacia el otro lado, etc. A la larga, tenemos un nudo bastante apretado. La probabilidad total es la distancia de un extremo al otro al cuadrado. *Casi toda aquella probabilidad acumulada se presenta en la región donde todas las flechas están en la misma dirección* (o en la misma fase). Todas las contribuciones de los caminos que tienen tiempos muy *diferentes* cuando cambiamos de camino, se anulan al apuntar en direcciones opuestas. Esto es porque, si escondemos las partes externas del espejo, todavía refleja casi exactamente lo mismo, porque todo lo que hicimos fue sacar un pedazo del diagrama dentro de los extremos de la espiral, y esto sólo produce un cambio muy pequeño en la luz. De manera que ésta es la relación entre la imagen última de los fotones con una probabilidad de llegada que depende de una acumulación de flechas y el principio del tiempo mínimo.

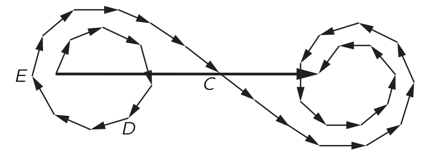


Figura 26-14. La suma de amplitudes de probabilidad para muchos caminos vecinos.

27

Óptica geométrica

27-1 Introducción

En este capítulo vamos a discutir algunas aplicaciones elementales de las ideas del capítulo anterior a algunos dispositivos prácticos, usando la aproximación llamada *óptica geométrica*. Es una aproximación muy útil en el diseño práctico de muchos sistemas e instrumentos ópticos. La óptica geométrica es o bien muy simple o bien muy complicada. Esto significa que la podemos estudiar sólo superficialmente, de manera que podamos diseñar instrumentos burdamente usando reglas que son tan simples que es escasamente necesario tratarlas aquí, ya que son prácticamente de nivel de enseñanza secundaria; o bien, si queremos saber de los pequeños errores en las lentes y detalles similares, el tema se hace tan complicado ¡que es muy avanzado para discutirlo aquí! Si uno tiene un problema real y detallado en el diseño de lentes, incluyendo el análisis de aberraciones, entonces se le aconseja leer sobre el tema, o bien simplemente trazar rayos a través de las diversas superficies (que es lo que el libro enseña cómo hacer), usando la ley de refracción de un lado a otro, y encontrar cómo salen y ver si forman una imagen satisfactoria. La gente ha dicho que esto es demasiado tedioso, pero hoy día la manera correcta de hacerlo es con computadores. Uno puede plantear el problema y hacer el cálculo para un rayo después de otro muy fácilmente. De manera que el tema es, en último término, realmente muy simple y no encierra nuevos principios. Además, resulta que las reglas, ya sea de la óptica elemental o de la avanzada, son rara vez características de otros campos, de manera que no hay ninguna razón especial para seguir más con el tema, con una excepción importante.

La teoría más avanzada y abstracta de la óptica geométrica fue desarrollada por Hamilton y resulta que ésta tiene aplicaciones muy importantes en mecánica. Es realmente aún más importante en mecánica que en óptica; así que dejamos la teoría de Hamilton para el tema de mecánica analítica avanzada que se estudia en cuarto año o en el posgrado. Así, estimando que la óptica geométrica contribuye muy poco, excepto por sí misma, vamos ahora a seguir discutiendo las propiedades elementales de sistemas ópticos simples sobre la base de los principios delineados en el capítulo anterior.

Para seguir, debemos tener una fórmula geométrica que es la siguiente: si tenemos un triángulo con una pequeña altura h y una base grande d, la diagonal s (la vamos a necesitar para encontrar la diferencia de tiempo entre dos caminos diferentes) es más larga que la base (figura 27-1). ¿Cuánto más larga? La diferencia $\Delta = s - d$ se puede encontrar de varias maneras. Una manera es ésta. Vemos que $s^2 - d^2 = h^2$, o $(s-d)(s+d) = h^2$. Pero $s - d = \Delta$ y $s + d \approx 2s$. Así,

$$\Delta \approx h^2/2s. \tag{27.1}$$

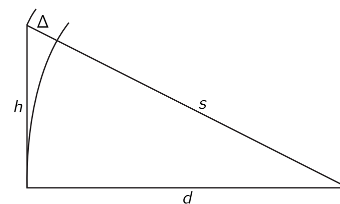


Figura 27-1

¡Ésta es toda la geometría que necesitamos para discutir la formación de imágenes por superficies curvas!

27-2 La distancia focal de una superficie esférica

La primera situación y la más simple a discutir es una superficie refractante única, que separa dos medios con diferentes índices de refracción (figura 27-2). Dejamos el caso de índices de refracción arbitrarios al estudiante, porque los *conceptos* son siempre lo más importante, no la situación específica, y el problema es suficientemente fácil de resolver en cualquier caso. Así que vamos a suponer

que a la izquierda la velocidad es 1 y a la derecha es $1/n$, donde n es el índice de refracción. La luz viaja más lentamente en el vidrio en un factor n.

Ahora supongan que tenemos un punto O, a una distancia s de la superficie frontal del vidrio, y otro punto O' a una distancia s' dentro del vidrio, y deseamos arreglar la superficie curva de tal manera que todo rayo desde O que toque la superficie en cualquier punto P se doble de tal manera que siga hacia el punto O'. Para que eso sea cierto, tenemos que darle forma a la superficie de tal manera que el tiempo que demora la luz en ir de O a P, o sea la distancia OP dividida entre la velocidad de la luz (la velocidad aquí es uno) más $n \cdot O'P$, que es el tiempo que demora en ir de P a O', sea igual a una constante independiente del punto P. Esta condición nos proporciona una ecuación para determinar la superficie. La respuesta es que la superficie es una curva de cuarto grado muy complicada, y el estudiante se puede entretener tratando de calcularla mediante la geometría analítica. Es más fácil tratar un caso especial que corresponde a $s \to \infty$, porque entonces la curva es de segundo grado y más reconocible. Es interesante comparar esta curva parabólica que encontramos para un espejo de enfoque cuando la luz viene del infinito.

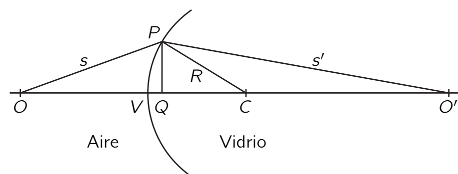


Figura 27-2. Enfoque mediante una superficie refractante única.

Así que la superficie apropiada no se puede hacer fácilmente; para enfocar la luz desde un punto a otro se necesita una superficie bastante complicada. Resulta en la práctica que no tratamos de hacer superficies tan complicadas ordinariamente, sino que en su lugar hacemos un compromiso. En vez de tratar de que *todos* los rayos lleguen al foco, nos arreglamos de manera que sólo los rayos bastante cerca del eje OO' vayan al foco. Los más lejanos se pueden desviar si quieren, desgraciadamente, porque la superficie ideal es complicada, y usamos en su lugar una superficie esférica con la curvatura correcta en el eje. Es tanto más fácil fabricar una esfera que otras superficies, que es útil para nosotros averiguar qué sucede a los rayos que chocan con una superficie esférica, suponiendo que sólo los rayos que pasan cerca del eje van a ser enfocados perfectamente. Estos rayos cerca del eje se llaman a veces *Rayos paraxiales*, y lo que estamos analizando son las condiciones para el enfoque de los rayos paraxiales. Vamos a discutir después los errores que se introducen por el hecho de que no todos los rayos están siempre cerca del eje.

Así, suponiendo que P está cerca del eje, bajamos una perpendicular PQ tal que la altura PQ sea h. Por un momento imaginamos que la superficie es un plano que pasa por P. En ese caso el tiempo necesario para ir de O a P va a exceder el tiempo de O a Q y también el tiempo de P a O' va a exceder el tiempo de Q a O'. Pero ésa es la razón de por qué el vidrio debe ser curvo, ¡porque el exceso de tiempo total debe ser compensado por el atraso al pasar de V a Q! Ahora bien, el *exceso* de tiempo a lo largo del camino OP es $h^2/2s$ y el exceso de tiempo en el otro camino es $nh^2/2s'$. Este tiempo de exceso, que debe ser igualado por el retraso al ir a lo largo de VQ, es diferente de lo que hubiera sido en el vacío, porque hay un medio presente. En otras palabras, el tiempo para ir de V a Q no es como si se fuera derecho en el aire, sino que es más lento en un factor n, de manera que el exceso de atraso en esta distancia es entonces $(n-1)VQ$. Y ahora, ¿qué largo tiene VQ? Si el punto C es el centro de la esfera y si su radio es R, vemos por la misma fórmula que la distancia VQ es igual a $h^2/2R$. Por lo tanto, descubrimos que la ley que conecta las distancias s y s' y que nos da el radio de curvatura R de la superficie que necesitamos es:

$$(h^2/2s) + (nh^2/2s') = (n-1)h^2/2R \tag{27.2}$$

o

$$(1/s) + (n/s') = (n-1)/R. \tag{27.3}$$

Si tenemos una posición O y otra posición O' y queremos enfocar la luz de O en O', podemos calcular el radio de curvatura R necesario mediante esta fórmula.

Ahora bien, resulta interesante que la misma lente, con la misma curvatura R, va a enfocar para otras distancias, a saber, para cualquier par de distancias tales que la suma de las dos inversas, una multiplicada por n, sea una constante. Así, una lente dada va (mientras nos limitemos a rayos paraxiales) a enfocar no sólo

de O a O', sino que entre un número infinito de otros pares de puntos, siempre que estos pares de puntos cumplan la relación que $1/s + n/s'$ sea una constante característica de la lente.

En particular, un caso interesante es cuando $s \to \infty$. Podemos ver en la fórmula que a medida que una s aumenta, la otra disminuye. En otras palabras, si el punto O se aleja, el punto O' se acerca y viceversa. A medida que el punto O va al infinito, el punto O' se sigue acercando hasta que alcanza una cierta distancia, llamada *distancia focal* f', dentro del material. Si llegan rayos paralelos, cortan el eje a una distancia f'. Igualmente podríamos imaginarlo al revés. (Recuerden la regla de la reciprocidad: si la luz va de O a O' también irá por supuesto de O' a O.) Por lo tanto, si tuviéramos una fuente de luz dentro del vidrio, nos gustaría saber dónde está el foco. Particularmente, si la luz en el vidrio estuviera en el infinito (el mismo problema) ¿dónde se enfocaría afuera? Esta distancia se llama f. Por supuesto que lo podemos plantear al revés. Si tuviéramos una fuente de luz en f y la luz pasara a través de la superficie, entonces seguiría como un rayo paralelo. Podemos encontrar fácilmente qué son f y f':

$$n/f' = (n-1)/R \qquad \text{o} \qquad f' = Rn/(n-1), \tag{27.4}$$

$$1/f = (n-1)/R \qquad \text{o} \qquad f = R/(n-1). \tag{27.5}$$

Vemos algo interesante: ¡Si dividimos cada distancia focal entre el índice de refracción correspondiente, obtenemos el mismo resultado! Este teorema, en efecto, es general. Es válido para cualquier sistema de lentes, por complicado que sea, de manera que es interesante recordarlo. Aquí no demostramos que fuera general —sencillamente lo hicimos ver para una superficie única, pero resulta ser válido en general que las dos distancias focales de un sistema están relacionadas de esta manera—. A veces la ecuación (27.3) se escribe de la siguiente manera:

$$1/s + n/s' = 1/f. \tag{27.6}$$

Ésta es más útil que (27.3), porque podemos medir f mucho más fácilmente que la curvatura y el índice de refracción de la lente: si no estamos interesados en diseñar una lente o en saber cómo se hizo, sino simplemente en sacarla del estante, la cantidad interesante es f ¡no la n, ni el 1 ni la R!

Ahora se presenta una situación interesante cuando s se hace menor que f. ¿Qué sucede entonces? Si $s < f$, entonces $(1/s) > (1/f)$ y, por lo tanto, s' es negativa; ¡nuestra ecuación dice que la luz se va a enfocar sólo para un valor negativo de s', no importa lo que eso significa! Significa algo muy interesante y bien definido. Todavía es una fórmula útil, en otras palabras, aun cuando los números sean negativos. Lo que significa se muestra en la figura 27-3. Si dibujamos los rayos que divergen de O, es cierto que se van a inclinar en la superficie, y no llegarán a un foco porque O está tan cerca que están “más allá de lo paralelo”. Sin embargo, divergen, como si hubieran venido de un punto O' *fuera* del vidrio. Ésta es una imagen aparente, a veces llamada *imagen virtual*. La imagen O' en la figura 27-2 se llama *imagen real*. Si la luz realmente llega a un punto, la imagen es real. Pero si la luz *parece* venir *desde* un punto, un punto ficticio, diferente del punto original, la imagen es virtual. De manera que cuando s' es negativa, significa que O' está al otro lado de la superficie y todo está bien.

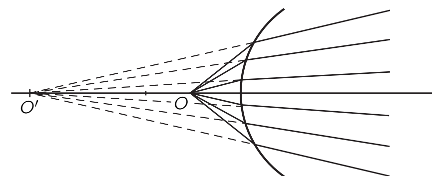


Figura 27-3. Una imagen virtual.

Ahora consideren el caso interesante cuando R es igual a infinito; entonces tenemos $(1/s) + (n/s') = 0$. En otras palabras $s' = -ns$, lo que significa que si miramos desde un medio denso a un medio menos denso y vemos un punto en el medio menos denso, parecerá estar más profundo en un factor n. Igualmente, podemos usar la misma ecuación al revés, de manera que si miramos hacia el interior de una superficie plana a un objeto que está a una cierta distancia dentro del medio denso, parecerá como si la luz no viniera de tan atrás (figura 27-4). Cuando miramos al fondo de una piscina desde arriba, no parece ser tan profunda como lo es realmente en un factor de 3/4, que es la inversa del índice de refracción del agua.

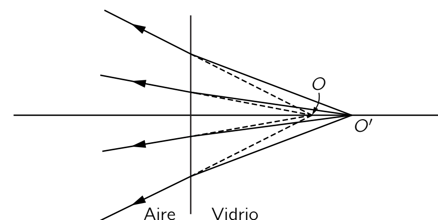


Figura 27-4. Una superficie plana corre la imagen de la luz desde O' hasta O.

Podríamos continuar, por supuesto, discutiendo el espejo esférico. Pero si alguien se da cuenta de las ideas puestas en juego debería ser capaz de desarrollarlo

solo. Por lo tanto, dejamos al estudiante que desarrolle la fórmula para el espejo esférico, pero mencionamos que está bien adoptar ciertas convenciones respecto a las distancias en juego:

(1) La distancia objeto s es positiva si el punto O está a la izquierda de la superficie.

(2) La distancia imagen s' es positiva si el punto O' está a la derecha de la superficie.

(3) El radio de curvatura de la superficie es positivo si el centro está a la derecha de la superficie.

En la figura 27-2, por ejemplo, s, s' y R son todas positivas; en la figura 27-3, s y R son positivas, pero s' es negativa. Si hubiéramos usado una superficie cóncava, nuestra fórmula (27.3) todavía daría el resultado correcto si hiciéramos simplemente R igual a una cantidad negativa.

Desarrollando la fórmula correspondiente para un espejo, usando las convenciones anteriores, encontrarían que si ponen $n = -1$ en toda la fórmula (27.3) (como si la sustancia detrás del espejo tuviera un índice -1) ¡se obtiene la fórmula correcta para un espejo!

Aunque la deducción de la fórmula (27.3) es simple y elegante, usando el tiempo mínimo, uno también puede, por supuesto, desarrollar la misma fórmula usando la ley de Snell, recordando que los ángulos son tan pequeños que los senos de los ángulos pueden ser remplazados por los ángulos mismos.

27-3 La distancia focal de una lente

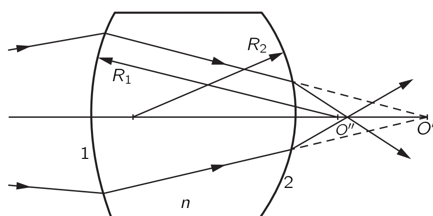


Figura 27-5. Formación de imagen por una lente de dos superficies.

Ahora continuemos considerando otra situación muy práctica. La mayoría de las lentes que usamos tienen dos superficies, no una solamente. ¿Cómo afecta esto las cosas? Supongan que tenemos dos superficies de diferente curvatura y el espacio entre ellas lleno de vidrio (figura 27-5). Queremos estudiar el problema de enfoque desde un punto O a otro punto O'. ¿Cómo lo podemos hacer? La respuesta es ésta: primero, usen la fórmula (27.3) para la primera superficie, olvidándose de la segunda superficie. Esto nos va a decir que la luz que divergía de O va a parecer convergiendo o divergiendo desde algún otro punto, digamos O', lo que depende del signo. Consideremos ahora un nuevo problema. Tenemos una superficie diferente, entre vidrio y aire, en la cual los rayos están convergiendo hacia un cierto punto O'. ¿Dónde van a converger realmente? ¡Usamos la misma fórmula de nuevo! Vemos que convergen en O''. ¡Así, si es necesario, podemos ir a través de 75 superficies simplemente usando la misma fórmula en sucesión, de una a la otra!

Hay algunas fórmulas de alta categoría que nos ahorrarían una energía considerable en las pocas veces que en nuestras vidas necesitáramos perseguir la luz a través de cinco superficies, pero es más fácil perseguirla a través de cinco superficies cuando surge el problema, que memorizar muchas fórmulas, porque ¡podría suceder que nunca tuviéramos que perseguirla a través de superficie alguna!

En todo caso, el principio es que cuando vamos a través de una superficie encontramos una nueva posición, un nuevo punto focal, y entonces tomamos ese punto como el punto de partida para la superficie siguiente y así sucesivamente. Para hacer esto realmente, ya que en la segunda superficie vamos de n a 1 en vez de 1 a n y ya que en muchos sistemas hay más de un tipo de vidrio, de manera que hay índices $n_1, n_2, \ldots$, realmente necesitamos una generalización de la fórmula (27.3) para el caso cuando hay dos índices diferentes n_1 y n_2 en vez de n solamente. Luego, no es difícil demostrar que la forma general de (27.3) es:

$$(n_1/s) + (n_2/s') = (n_2 - n_1)/R. \tag{27.7}$$

Particularmente simple es el caso en el cual las dos superficies están muy juntas —tan juntas que podemos ignorar pequeños errores debido al grosor—. Si

dibujamos la lente como se muestra en la figura 27-6, nos podemos hacer esta pregunta: ¿Cómo debe construirse la lente para que enfoque la luz de O a O'? Supongan que la luz llega exactamente al borde de la lente en el punto P. Luego, el exceso de tiempo en ir de O a O' es $(n_1h^2/2s) + (n_1h^2/2s')$, ignorando por un instante la presencia del grosor T del vidrio de índice n_2. Ahora bien, para hacer el tiempo del camino directo igual al del camino OPO', tenemos que usar un pedazo de vidrio cuyo grosor T en el centro sea tal que la demora introducida al ir a través de este grosor sea suficiente para compensar el exceso de tiempo indicado. Luego, el grosor de la lente en el centro debe estar dado por la relación

$$(n_1h^2/2s) + (n_1h^2/2s') = (n_2 - n_1)T. \tag{27.8}$$

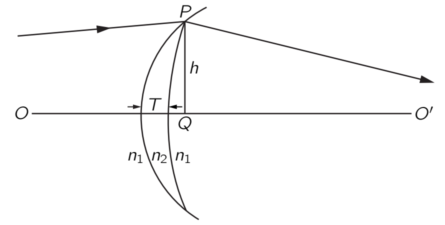


Figura 27-6. Lente delgada con dos radios positivos.

También podemos expresar T términos de los radios R_1 y R_2 de las dos superficies. Poniendo atención a nuestra convención (3) encontramos así, para $R_1 < R_2$ (lente convexa),

$$T = (h^2/2R_1) - (h^2/2R_2). \tag{27.9}$$

Por lo tanto, obtenemos finalmente

$$(n_1/s) + (n_1/s') = (n_2 - n_1)(1/R_1 - 1/R_2). \tag{27.10}$$

Notamos de nuevo que si uno de los puntos está en el infinito, el otro va a estar en un punto que llamamos distancia focal f. La distancia focal f está dada por

$$1/f = (n - 1)(1/R_1 - 1/R_2). \tag{27.11}$$

donde $n = n_2/n_1$.

Ahora, si tomamos el caso opuesto, que s vaya a infinito, vemos que s' está a la distancia focal f'. Esta vez las distancias focales son iguales. (Éste es otro caso especial de la regla general que el cociente entre las dos distancias focales es el cociente entre los índices de refracción de los dos medios donde los rayos se enfocan. En este sistema óptico particular, los índices inicial y final son los mismos, de manera que las dos distancias focales son iguales.)

Olvidando por un momento la fórmula real para la distancia focal, si compráramos una lente que alguien diseñó con ciertos radios de curvatura y cierto índice, podríamos medir la distancia focal, digamos, viendo dónde se enfoca un punto en el infinito. Una vez obtenida la distancia focal, sería mejor escribir nuestra ecuación, directamente en función de la distancia focal y entonces la fórmula es

$$(1/s) + (1/s') = 1/f. \tag{27.12}$$

Veamos ahora cómo funciona la fórmula y lo que implica en diferentes circunstancias. Primero, implica que si s o s' es infinita, la otra es f. Esto significa que la luz paralela se enfoca a una distancia f y esto, en efecto, *define a f*. Otra cosa interesante que dice es que ambos puntos se mueven en la misma dirección. Si uno se mueve hacia la derecha el otro también lo hace. Otra cosa que dice es que s y s' son iguales si ambos son iguales a $2f$. En otras palabras, si queremos una situación simétrica encontramos que ambas se enfocan a una distancia $2f$.

27-4 Aumento

Hasta aquí, discutimos sólo el enfoque para puntos en el eje. Discutamos ahora la formación de imágenes de objetos que no están en el eje exactamente, sino que un poco corrido, de manera que podamos entender las propiedades del *aumento*. Cuando disponemos de una lente para enfocar la luz proveniente de un pequeño filamento en un "punto" de una pantalla, encontramos que en la pantalla obtenemos una "fotografía" del mismo filamento excepto que es de tamaño mayor o menor que el filamento verdadero. Esto debe significar que la luz llega a un foco desde *cada punto* del filamento. Para entender esto un poco mejor, analicemos el sistema de lente delgada que se muestra esquemáticamente en la figura 27-7. Conocemos los siguientes hechos:

(1) Cualquier rayo que llegue paralelo por un lado prosigue hasta un cierto punto particular llamado foco en el otro lado a una distancia de la lente.

(2) Cualquier rayo que llega a la lente desde el foco de un lado sale paralelo al eje por el otro lado.

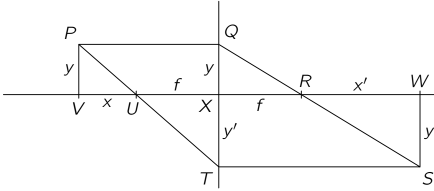


Figura 27-7. Geometría de la producción de imagen por una lente delgada.

Esto es todo lo que necesitamos para establecer la fórmula (27.12) por geometría, como sigue: supongan que tenemos un objeto a cierta distancia x del foco, sea y la altura del objeto. Entonces sabemos que uno de los rayos, a saber PQ, se va a inclinar de manera de pasar por el foco R en el otro lado. Ahora bien, si la lente va a enfocar el punto P de cualquier manera, podemos averiguar dónde, si averiguamos por dónde pasa un rayo más solamente, porque el nuevo foco va a estar donde los dos se vuelven a cortar. Necesitamos sólo usar nuestro ingenio para encontrar la dirección exacta de *algún* otro rayo. Pero nos acordamos de que un rayo paralelo pasa por el foco y *viceversa*: ¡un rayo que pasa por el foco va a salir paralelo! Así que dibujamos el rayo PT pasando por U. (Es cierto que los verdaderos rayos que están haciendo el enfoque pueden ser mucho más limitados que los dos que hemos dibujado, pero son más difíciles de imaginar, así hacemos creer que podemos hacer este rayo.) Ya que saldría paralelo dibujamos TS paralelo a XW. La intersección S es el punto que necesitamos. Esto va a determinar el lugar correcto y la altura correcta. Llamemos la altura y' y la distancia al foco x'. Ahora podemos deducir una fórmula de la lente. Usando los triángulos semejantes PVU y TXU, encontramos

$$\frac{y'}{f} = \frac{y}{x}. \tag{27.13}$$

Igualmente, de los triángulos SWR y QXR, obtenemos

$$\frac{y'}{x'} = \frac{y}{f}. \tag{27.14}$$

Despejando y'/y de cada una encontramos que

$$\frac{y'}{y} = \frac{x'}{f} = \frac{f}{x}. \tag{27.15}$$

La ecuación (27.15) es la famosa fórmula de la lente; en ella está todo lo que necesitamos saber de las lentes: nos da el aumento y'/y en términos de las distancias y las distancias focales. También relaciona las dos distancias x y x' con f:

$$xx' = f^2, \tag{27.16}$$

que tiene una forma más sencilla para trabajar con ella que la ecuación (27.12). Dejamos al estudiante que demuestre que si llamamos $s = x + f$ y $s' = x' + f$, la ecuación (27.12) es lo mismo que la ecuación (27.16).

27-5 Lentes compuestas

Sin deducirlo realmente, vamos a describir en forma breve el resultado general cuando tenemos varias lentes. Si tenemos un sistema de varias lentes, ¿cómo es posible analizarlo? Es fácil. Empezamos con algún objeto y calculamos dónde está su imagen para la primera lente usando la fórmula (27.16) o (27.12) o cualquier otra fórmula equivalente o dibujando diagramas. Así, encontramos una imagen. Después tratamos esta imagen como la fuente para la lente siguiente y usamos la segunda lente, sea cualquiera su distancia focal, para encontrar de nuevo una imagen. Simplemente perseguimos la cosa a través de la sucesión de lentes. Y eso es todo lo que hay que hacer. No encierra nada nuevo en principio; así que no vamos a entrar en detalles. Sin embargo, hay un resultado neto muy interesante de los efectos de cualquier sucesión de lentes acerca de la luz que parte y termina en el mismo medio, aire, digamos. Cualquier instrumento óptico —un telescopio o un microscopio con una serie de lentes y espejos— tiene la siguiente

propiedad: existen dos planos llamados *planos principales* del sistema (estos planos están a menudo muy cerca de la primera superficie de la primera lente y de la última superficie de la última lente), que tienen las siguientes propiedades: (1) Si la luz llega al sistema paralela por el primer lado, llega a un cierto foco a una distancia del *segundo* plano principal igual a la distancia focal, exactamente como si el sistema fuera una lente delgada situada en este plano. (2) Si la luz llega paralela por el otro lado, llega a un foco a la misma distancia f del *primer* plano principal, de nuevo como si hubiera una lente delgada colocada ahí. (Vean la figura 27-8.)

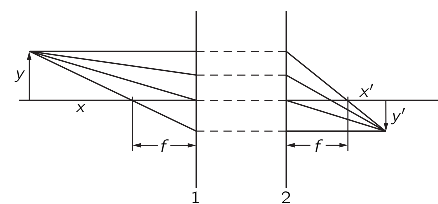


Figura 27-8. Ilustración de los planos principales de un sistema óptico.

Por supuesto, si medimos las distancias x y x', y y y y' como antes, la fórmula (27.16) que hemos escrito para una lente delgada es absolutamente general, siempre que midamos la distancia focal desde los planos principales y no desde el centro de la lente. Sucede que para una lente delgada los planos principales son coincidentes. Es como si pudiéramos tomar una lente delgada, rebanarla por la mitad y separarla y no nos diéramos cuenta de que se había separado. ¡Cada rayo que llega salta inmediatamente por el otro lado del segundo plano por el mismo punto que entró en el primer plano! Los planos principales y la distancia focal pueden ser encontrados, ya sea por experimento o por cálculo y entonces todo el conjunto de propiedades del sistema óptico queda descrito. Es muy interesante que el resultado no es complicado cuando hemos terminado con un sistema óptico tan grande y complicado.

27-6 Aberraciones

Antes de que nos estusiasmemos mucho con lo maravillosas que son las lentes, debemos apresurarnos en agregar que también hay serias limitaciones, debido a que nos hemos limitado estrictamente hablando a rayos paraxiales, los rayos cerca del eje. Una lente real con tamaño finito presentará, en general, *aberraciones*. Por ejemplo, un rayo que está en el eje por supuesto que pasa por el foco. Un rayo que está muy cerca del eje todavía pasará muy bien por el foco. Pero, a medida que nos alejamos, el rayo se empieza a desviar del foco, quizá porque queda corto, y un rayo que da con el borde superior baja y pierde el foco por una cantidad bastante grande. Así, en vez de obtener una imagen puntual, obtenemos un manchón. Este efecto se llama *aberración esféricass* porque es una propiedad de las superficies esféricas que usamos en vez de la de forma correcta. Esto podría ser remediado para cualquier distancia del objeto específica, cambiando la forma de la superficie de la lente o tal vez usando varias lentes arregladas de modo que las aberraciones de las lentes individuales tienden a cancelarse.

Las lentes tienen otro defecto: la luz de diferentes colores tiene velocidades diferentes o índices de refracción diferentes en el vidrio, y por lo tanto la distancia focal de una lente dada es diferente para distintos colores. De manera que si imaginamos un punto blanco, la imagen va a tener colores, porque cuando enfocamos para el rojo, el azul está fuera de foco o viceversa. Esta propiedad se llama *aberración cromática*.

Hay todavía otros defectos. Si el objeto está fuera del eje, entonces realmente el foco ya no es perfecto, cuando aquel se aleja bastante del eje. La manera más fácil de verificarlo es enfocar una lente y después inclinarla de manera que los rayos lleguen formando un ángulo grande con el eje. Entonces, la imagen que se forme será generalmente muy aproximada y puede que no haya ningún lugar donde se enfoque bien. Hay así varios tipos de errores en las lentes que el diseñador óptico trata de remediar usando muchas lentes para compensar los errores de unas con los de las otras.

¿Cuánto cuidado debemos tener para eliminar las aberraciones? ¿Es posible hacer un sistema óptico perfecto? Supongan que hemos construido un sistema óptico que se supone que va a traer la luz exactamente a un punto. Ahora bien, razonando desde el punto de vista del tiempo mínimo, ¿podemos encontrar una condición que indique qué tan perfecto debe ser el sistema? El sistema va a tener cierto tipo de abertura de entrada para la luz. Si tomamos el rayo más alejado del eje que puede llegar al foco (si el sistema es perfecto, por supuesto),

los tiempos para todos los rayos son exactamente iguales. Pero nada es perfecto, de manera que la pregunta es ¿hasta qué punto puede ser errado el tiempo para este rayo tal que no valga la pena hacer una corrección mayor? Eso depende de lo perfecta que queramos hacer la imagen. Pero supongan que queremos hacer la imagen tan perfecta como se pueda. Entonces nuestra impresión, por supuesto, es que tenemos que arreglar que cada rayo se demore un tiempo lo más parecido posible. Pero resulta que esto no es cierto, que más allá de un cierto punto estamos tratando de hacer algo que es demasiado fino ¡porque la teoría de la óptica geométrica no funciona!

Recuerden que el principio del tiempo mínimo no es una formulación precisa, a diferencia del principio de la conservación de la energía o del principio de conservación del *momentum*. El principio del tiempo mínimo es sólo una *aproximación* y es interesante saber cuánto error puede permitirse y todavía no producir una diferencia aparente. La respuesta es que si hemos dispuesto que entre el rayo máximo —el peor rayo, el rayo que está más lejano— y el rayo central, la diferencia de tiempo sea menor que más o menos el periodo que corresponde a una oscilación de la luz, entonces es inútil un mayor mejoramiento. La luz es algo oscilatorio con una frecuencia definida que está relacionada con la longitud de onda y, si hemos dispuesto que la diferencia de tiempo para rayos diferentes sea menor que un periodo más o menos, es inútil ir más allá.

27-7 Poder de resolución

Otra pregunta interesante —una pregunta teórica muy importante para todos los instrumentos ópticos— es qué *poder de resolución* tienen. Si construimos un microscopio, queremos ver los objetos que estamos mirando. Eso significa, por ejemplo, que si estamos mirando una bacteria con un punto a cada lado, queremos *ver* que hay dos puntos cuando los aumentamos. Uno podría pensar que todo lo que tenemos que hacer es obtener suficiente aumento —siempre podemos agregar otra lente y siempre podemos aumentar una y otra vez y, con la inteligencia de los diseñadores, todas las aberraciones esféricas y aberraciones cromáticas pueden ser anuladas y no existe razón alguna para que no podamos seguir aumentando la imagen—. De manera que las limitaciones de un microscopio no son que sea imposible construir una lente que aumente más de 2 000 diámetros. Podemos construir un sistema de lentes que aumenta 10 000 diámetros, pero *aún* podríamos no ver dos puntos que están muy juntos debido a las limitaciones de la óptica geométrica, debido a que el tiempo mínimo no es preciso.

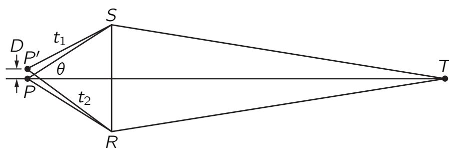


Figura 27-9. Poder de resolución de un sistema óptico.

Para descubrir la regla que determina a qué distancia deben estar dos puntos de manera que en la imagen aparezcan como puntos separados, puede establecerse un procedimiento muy hermoso en relación con el tiempo que demoran los diferentes rayos. Supongan que despreciamos las aberraciones ahora e imaginen que para un cierto punto P (figura 27-9) todos los rayos desde el objeto a la imagen T demoran exactamente el mismo tiempo. (Esto no es cierto, porque no es un sistema perfecto, pero eso es otro problema.) Ahora tomen otro punto cercano P' y pregunten si su imagen va a ser distinta de T. En otras palabras, si podemos notar la diferencia entre ellos. Por supuesto, de acuerdo con la óptica geométrica, debería haber dos puntos imagen, pero lo que vemos puede ser bastante borroso y podríamos no ser capaces de darnos cuenta que hay dos puntos. La condición para que el segundo punto sea enfocado en un lugar claramente diferente del primero, es que los dos tiempos para ir de un extremo a otro de los rayos extremos $P'ST$ y $P'RT$ a cada lado de la abertura grande de las lentes, *no* deben ser iguales desde los dos puntos objeto posibles hasta un punto imagen dado. ¿Por qué? Porque si los tiempos fueran iguales, por supuesto, ambos se *enfocarían* en el mismo punto. De manera que los tiempos no serán iguales. Pero ¿en cuánto tienen que diferir para que podamos decir que ambos *no* llegan a un foco común, de manera que podamos distinguir los dos puntos imagen? La regla general para la resolución de cualquier instrumento óptico es ésta: dos fuentes puntuales diferentes pueden ser resueltas sólo si una fuente es enfocada en un punto tal que los tiempos en que los rayos máximos de la otra fuente lleguen a ese punto, comparado con su punto

imagen propio, difieran en más de un periodo. Es necesario que la diferencia de tiempo entre el rayo superior y el inferior al foco *equivocado* exceda una cierta cantidad, a saber, aproximadamente el periodo de oscilación de la luz:

$$t_2 - t_1 > 1/\nu, \tag{27.17}$$

donde v es la frecuencia de la luz (número de oscilaciones por segundo; también la velocidad dividida entre la longitud de onda). Si la distancia de separación de los dos puntos se llama D y si el ángulo de abertura de la lente se llama θ, uno puede demostrar que (27.17) es exactamente equivalente a la afirmación de que D debe exceder $\lambda/n \operatorname{sen}\theta$, donde n es el índice de refracción en P y λ es la longitud de onda. Las cosas más pequeñas que podemos ver son por lo tanto aproximadamente la longitud de onda de la luz. Una fórmula correspondiente existe para los telescopios, que nos da la menor diferencia angular entre dos estrellas que justo alcance a distinguirse.*

* El ángulo es alrededor de λ/D, donde D es el diámetro de la lente. ¿Pueden ver por qué?

28

Radiación electromagnética

28-1 Electromagnetismo

Los momentos más espectaculares en el desarrollo de la física son aquellos en los cuales se producen grandes síntesis, donde se descubre súbitamente que fenómenos que antes habían parecido diferentes no son sino distintos aspectos de la misma cosa. La historia de la física es la historia de tales síntesis y la base del éxito de la ciencia física está principalmente en que somos *capaces* de sintetizar.

Quizás el momento más espectacular en el desarrollo de la física durante el siglo XIX se le presentó a J. C. Maxwell un día alrededor de 1860 cuando combinó las leyes de la electricidad y del magnetismo con las leyes del comportamiento de la luz. Como resultado, las propiedades de la luz fueron desenredadas parcialmente —esa antigua y sutil materia que es tan importante y misteriosa que se creyó necesario arreglar una creación especial para ella al escribir el Génesis—. Maxwell pudo decir, al terminar su descubrimiento: "¡Que haya electricidad y magnetismo y allí está la luz!"

Para este momento culminante hubo una larga preparación en el descubrimiento y el desarrollo gradual de las leyes de la electricidad y del magnetismo. Esta historia la vamos a reservar para un estudio detallado el próximo año. Sin embargo, la historia es, brevemente, como sigue. Las propiedades de la electricidad y del magnetismo, de las fuerzas eléctricas de atracción y repulsión, y de las fuerzas magnéticas descubiertas gradualmente, mostraron que, aunque estas fuerzas eran bastante complejas, todas disminuían con el cuadrado de la distancia. Sabemos, por ejemplo, que la sencilla ley de Coulomb para cargas estáticas es que el campo de fuerza eléctrica varía inversamente con el cuadrado de la distancia. Como consecuencia, para distancias suficientemente grandes hay poca influencia de un sistema de cargas sobre otro. Maxwell notó que las ecuaciones o leyes que se habían descubierto hasta ese momento eran incompatibles entre sí cuando trató de juntarlas, y para que todo el sistema fuera compatible tuvo que agregar otro término a sus ecuaciones. Con este nuevo término surgió una predicción asombrosa que fue que una parte de los campos eléctricos y magnéticos disminuiría mucho más lentamente con la distancia que la inversa del cuadrado, a saber, ¡inversamente con la primera potencia de la distancia! y así se dio cuenta de que las corrientes eléctricas en un lugar pueden afectar a otras cargas lejanas y predijo los efectos básicos con los que estamos familiarizados hoy día —transmisión de radio, radar, etcétera—.

Parece un milagro que alguien hablando en Europa pueda, con simples influencias eléctricas, ser oído a miles de kilómetros en Los Ángeles. ¿Cómo es posible? Lo es porque los campos no varían con la inversa del cuadrado, sino que únicamente con la inversa de la primera potencia de la distancia. Finalmente, entonces, se reconoció que incluso la luz consistía en influencias eléctricas y magnéticas, que se extienden sobre grandes distancias, generadas por una oscilación increíblemente rápida de los electrones en los átomos. Todos estos fenómenos los resumimos mediante la palabra *radiación* o más específicamente *radiación electromagnética*, habiendo también uno o dos tipos más de radiación. Casi siempre radiación significa radiación electromagnética.

Y así está enlazado el universo. Los movimientos atómicos de una estrella distante todavía tienen suficiente influencia a esta gran distancia para poner los electrones de nuestro ojo en movimiento y así sabemos de las estrellas. ¡Si esta ley no existiera estaríamos literalmente a oscuras respecto al mundo exterior! Y los oleajes eléctricos en una galaxia distante 5 000 millones de años luz —que

es el objeto más lejano que hemos encontrado hasta ahora— pueden influenciar todavía de una manera significativa y detectable las corrientes en el gran "plato" frente a un radiotelescopio. Y así es cómo vemos las estrellas y las galaxias.

Este notable fenómeno es lo que vamos a discutir en el presente capítulo. Al comienzo de este curso de física delineamos un amplio cuadro del mundo, pero ahora estamos más preparados para entender algunos aspectos de él y así vamos a volver ahora sobre algunos puntos con mayor detalle. Empezamos describiendo la posición de la física al final del siglo XIX. Todo lo que se conocía entonces de las leyes fundamentales se puede resumir como sigue.

Primero, hubo leyes de fuerzas: una fuerza estaba dada por la ley de gravitación que hemos escrito varias veces; la fuerza sobre un objeto de masa m debida a otra masa M está dada por

$$\boldsymbol{F} = GmM\boldsymbol{e}_r/r^2, \tag{28.1}$$

donde $\boldsymbol{e}_r$ es un vector dirigido de m a M y r es la distancia entre ellas.

A continuación, las leyes de la electricidad y del magnetismo, como se conocían al final del siglo XIX, son éstas: las fuerzas eléctricas que actúan sobre una carga q pueden ser descritas mediante dos campos, llamados $\boldsymbol{E}$ y $\boldsymbol{B}$, y la velocidad $\boldsymbol{v}$ de la carga q mediante la ecuación

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \tag{28.2}$$

Para completar esta ley, tenemos que decir cuáles son las fórmulas para $\boldsymbol{E}$ y $\boldsymbol{B}$ en una circunstancia dada: si varias cargas están presentes, $\boldsymbol{E}$ y $\boldsymbol{B}$ son cada una la suma de contribuciones, una por cada carga individual. ¡De manera que si podemos encontrar el $\boldsymbol{E}$ y el $\boldsymbol{B}$ producido por una carga sola necesitamos solamente sumar todos los efectos de todas las cargas del universo para obtener el $\boldsymbol{E}$ y el $\boldsymbol{B}$ total! Éste es el principio de superposición.

¿Cuál es la fórmula para el campo eléctrico y magnético producido por una carga individual? Resulta que esto es muy complicado y se necesita mucho estudio y muchos refinamientos para apreciarlo. Pero ése no es el punto. Escribimos ahora la ley sólo para impresionar al lector con la belleza de la naturaleza, por así decirlo, o sea, que es *posible* resumir todo el conocimiento fundamental en una página con notaciones con las cuales ahora está familiarizado. Esta ley para los campos de una carga individual es completa y exacta, hasta donde sabemos (con excepción de la mecánica cuántica), pero se ve bastante complicada. No vamos a estudiar todas las partes ahora, sólo la escribimos para dar una impresión, mostrar que se puede escribir y de manera que podamos ver con anticipación cómo se ve aproximadamente. De hecho, la manera más *útil* de escribir las leyes correctas de la electricidad y del magnetismo no es como las vamos a escribir ahora, pero encierra lo que se llama *ecuaciones de campo*, que vamos a aprender el próximo año. Pero las notaciones matemáticas para éstas son diferentes y nuevas. Así que escribimos la ley en una forma inconveniente para los cálculos, pero en una notación que conocemos ahora.

El campo eléctrico $\boldsymbol{E}$ está dado por

$$\boldsymbol{E} = \frac{-q}{4\pi\epsilon_0}\left[\frac{\boldsymbol{e}_{r'}}{r'^2} + \frac{r'}{c}\frac{d}{dt}\left(\frac{\boldsymbol{e}_{r'}}{r'^2}\right) + \frac{1}{c^2}\frac{d^2}{dt^2}\boldsymbol{e}_{r'}\right]. \tag{28.3}$$

¿Qué nos dicen los diversos términos? Tomen el primero, $\boldsymbol{E} = -q\boldsymbol{e}_{r'}/4\pi\epsilon_0 r'^2$. Es, por supuesto, la ley de Coulomb que ya conocemos: q es la carga que está produciendo el campo; $\boldsymbol{e}_{r'}$ es el vector en la dirección desde el punto P donde se mide $\boldsymbol{E}$, r la distancia de P a q. Pero, la ley de Coulomb está equivocada. Los descubrimientos del siglo XIX demostraron que las influencias no pueden viajar más rápido que una cierta velocidad fundamental c, que ahora llamamos velocidad de la luz. No es correcto que el primer término sea la ley de Coulomb, no sólo porque no es posible conocer dónde está la carga *ahora* y a qué distancia está *ahora*, sino también porque lo único que puede afectar el campo en un lugar y en un tiempo dados es el comportamiento de las cargas en el *pasado*. ¿Qué tan

lejos en el pasado? La tardanza en el tiempo, o el así llamado *tiempo retardado*, es el tiempo que se necesita a la velocidad de la luz para llegar desde la carga hasta el punto P del campo. El atraso es r'/c.

Así, para permitir este atraso en el tiempo ponemos una pequeña comilla en r significando lo lejos que *estaba* cuando la información que ahora llega a P dejó a q. Sólo por un momento supongamos que la carga llevará una luz y que la luz pudiera llegar solamente con la velocidad c. Entonces, cuando miramos a q, no veríamos donde está ahora, por supuesto, sino dónde *estaba* en cierto tiempo anterior. Lo que aparece en nuestra fórmula es la dirección *aparente* $\theta_{r'}$ —la dirección que era—, la así llamada dirección *retardada* —y a la distancia *retardada* r'—. Esto sería suficientemente fácil de entender, pero también está equivocado. Todo es mucho más complicado.

Hay varios términos más. El término siguiente es como si la naturaleza estuviera tratando de tomar en cuenta el hecho de que el efecto es retardado, si podemos decirlo tan burdamente. Sugiere que deberíamos calcular el campo de Coulomb retardado y agregarle una corrección que es su rapidez de variación por el tiempo retardado que usamos. Parece que la naturaleza estuviera tratando de adivinar lo que va a ser el campo en este momento tomando la rapidez de variación y multiplicando por el tiempo en que está retardado. Pero no hemos terminado aún. Hay un tercer término —la segunda derivada, respecto a t del vector en la dirección de la carga—. Ahora, la fórmula *está* completa y eso es todo lo que hay con relación al campo eléctrico de una carga que se mueve arbitrariamente.

El campo magnético está dado por:

$$\boldsymbol{B} = -\boldsymbol{e}_{r'} \times \boldsymbol{E}/c. \tag{28.4}$$

Hemos escrito esto sólo con el propósito de mostrar la belleza de la naturaleza o, en cierto sentido, el poder de la matemática. No pretendemos entender *por qué* es posible escribir tanto en un espacio tan pequeño, pero (28.3) y (28.4) contienen el mecanismo según el cual funcionan los generadores eléctricos, cómo funciona la luz, todos los fenómenos de la electricidad y el magnetismo. Para completar la historia también necesitamos conocer, por supuesto, algo acerca del comportamiento de los materiales que intervienen —las propiedades de la materia— que no están descritas en forma apropiada por (28.3).

Para terminar nuestra descripción del mundo del siglo XIX debemos mencionar otra gran síntesis que ocurrió en ese siglo, una con la cual Maxwell también tuvo mucho que ver: fue la síntesis de los fenómenos del calor y la mecánica. Estudiaremos pronto ese tema.

Lo que debió agregarse en el siglo XX fue que se encontró que todas las leyes dinámicas de Newton estaban equivocadas y hubo que introducir la mecánica cuántica para corregirlas. Las leyes de Newton son técnicamente válidas cuando la escala de las cosas es suficientemente grande. Estas leyes cuánticas, combinadas con las leyes de la electricidad, sólo recientemente han sido combinadas para formar un conjunto de leyes llamado *electrodinámica cuántica*. Además, se descubrieron varios fenómenos nuevos de los cuales el primero fue la radiactividad, descubierta por Antoine Henri Becquerel en 1896 —justo alcanzó a meterla de contrabando en el siglo XIX—. Se investigó este fenómeno de la radiactividad hasta producir nuestro conocimiento de los núcleos y nuevos tipos de fuerzas que no son gravitacionales ni eléctricas, sino nuevas partículas con diferentes interacciones, un tema que aún no ha sido desenredado.

Para aquellos puristas que saben más (los profesores que podrían estar leyendo esto) deberíamos agregar que cuando decimos que (28.3) es una expresión completa del conocimiento de la electrodinámica, no estamos siendo completamente rigurosos. Hubo un problema que no fue completamente resuelto a fines del siglo XIX. Cuando tratamos de calcular el campo de todas las cargas *incluyendo la carga misma sobre la cual queremos que actúe el campo*, encontramos dificultades tratando de encontrar la distancia, por ejemplo, de una carga respecto a sí misma y dividiendo algo entre esa distancia que es cero. El problema de cómo manejar la

parte de este campo que es generada por la misma carga en la cual queremos que actúe el campo no se ha resuelto todavía. Así que lo dejamos ahí; no tenemos aún una solución completa de ese enigma y, por lo tanto, vamos a evitar el enigma mientras podamos.

28-2 Radiación

Ése es entonces un resumen de la imagen del mundo. Usémosla ahora para discutir los fenómenos llamados radiación. Para discutir estos fenómenos, debemos seleccionar de la ecuación (28.3) sólo la parte que varía inversamente con la distancia y no con el cuadrado de la distancia. Resulta que cuando finalmente encontramos esa parte, es tan simple en su forma que es legítimo estudiar óptica y electrodinámica de una manera elemental tomándola como "la ley" del campo eléctrico producido por una carga lejana en movimiento. Vamos a considerarla temporalmente como una ley dada que vamos a aprender en detalle el próximo año.

De los términos que aparecen en (28.3) el primero evidentemente varía inversamente con el cuadrado de la distancia y el segundo es sólo una corrección por el retardo; así que es fácil demostrar que ambos varían inversamente con el cuadrado de la distancia. Todos los efectos en los que estamos interesados provienen del tercer término, que no es muy complicado, después de todo. Lo que este término dice es: miren la carga y observen la dirección del vector (podemos proyectar su extremo sobre la superficie de una esfera unitaria). A medida que la carga se mueve, el vector se mueve rápidamente y es *la aceleración de ese vector lo que estamos buscando.* Eso es todo. Así,

$$\boldsymbol{E} = \frac{-q}{4\pi\epsilon_0 c^2}\,\frac{d^2\boldsymbol{e}_{r'}}{dt^2} \tag{28.5}$$

es una expresión de las leyes de la radiación, porque éste es el único término importante cuando nos alejamos lo suficiente para que los campos varíen inversamente con la distancia; las partes que varían con el cuadrado han disminuido tanto que no estamos interesados en ellas.

Ahora podemos ir un poco más allá en el estudio de (28.5) para ver qué significa. Supongan que una carga se está moviendo de cualquier manera y que la estamos observando desde cierta distancia. Imaginemos por un momento que en un cierto sentido está "encendida" (a pesar de que es la luz lo que estamos tratando de explicar), la imaginamos como un pequeño punto blanco. Entonces veríamos este punto blanco corriendo de un lado a otro. Pero no vemos *exactamente* cómo está corriendo *ahora* mismo debido al atraso del que hemos estado hablando. Lo que importa es cómo se estaba moviendo *antes.* El vector $\boldsymbol{e}_{r'}$ apunta hacia la posición aparente de la carga. El extremo de $\boldsymbol{e}_{r'}$, por supuesto, se mueve sobre una pequeña curva tal que su aceleración tenga dos componentes. Una es la parte transversal, porque su extremo va hacia arriba y hacia abajo y la otra es una parte radial porque se mantiene sobre una esfera. Es fácil demostrar que la última es mucho menor y que varía inversamente con el cuadrado de r, cuando r es muy grande. Esto es fácil de ver porque cuando imaginamos que alejamos cada vez más una fuente dada, entonces los movimientos de él se ven cada vez más chicos, inversamente con la distancia, pero la componente radial de la aceleración está variando mucho más rápido que con la inversa de la distancia. Así que para fines prácticos todo lo que tenemos que hacer es proyectar el movimiento en un plano a distancia unitaria. Por lo tanto, encontramos la siguiente regla: imaginen que miramos a la carga en movimiento y que todo lo que vemos está atrasado —como un pintor tratando de pintar una escena en una pantalla a una distancia unitaria—. Un pintor real, por supuesto, no toma en cuenta el hecho que la luz vaya con una cierta velocidad, sino que pinta el mundo como lo ve. Deseamos ver cómo luciría su cuadro. De manera que vemos un punto, que representa la carga moviéndose en el cuadro. La aceleración de ese punto es proporcional al campo eléctrico. Eso es todo —todo lo que necesitamos—.

Así pues, la ecuación (28.5) es la fórmula completa y correcta para la radiación; aun los efectos relativistas están incluidos. Sin embargo, a menudo queremos aplicarla a circunstancias aún más simples en las cuales las cargas se mueven sólo una pequeña distancia con una velocidad relativamente baja. Ya que se están moviendo lentamente, no se mueven una distancia muy apreciable desde donde parten, de manera que el tiempo de retraso es prácticamente constante. Entonces la ley es aún más simple, porque el atraso en el tiempo es fijo. Así, imaginamos que la carga está ejecutando un movimiento muy pequeño a una distancia efectivamente constante. El atraso a la distancia r es r/c. Entonces nuestra regla se convierte en la siguiente: si el objeto cargado se está moviendo con un movimiento muy pequeño y se desplaza lateralmente, la distancia $x(t)$, el ángulo en que el vector $\boldsymbol{e}_{r'}$ se desplaza es x/r y como r es prácticamente constante, la componente x de $d^2\boldsymbol{e}_{r'}/dt^2$ es simplemente la aceleración de x en un tiempo anterior dividido entre r, de manera que finalmente obtenemos la ley que queremos, que es:

$$E_x(t) = \frac{-q}{4\pi\epsilon_0 c^2 r}\, a_x\Big(t - \frac{r}{c}\Big). \tag{28.6}$$

Sólo la componente a_x, perpendicular a la visual, es importante. Veamos por qué. Evidentemente, si la carga se está moviendo hacia y desde nosotros, el vector en esa dirección no cambia y no tiene aceleración. Así, pues, sólo el movimiento transversal es importante, sólo la aceleración que vemos proyectada en la pantalla.

28-3 El radiador dipolar

Como nuestra "ley" fundamental de la radiación electromagnética, vamos a suponer la validez de la ecuación (28.6), es decir, que el campo eléctrico producido por una carga acelerada que se está moviendo en forma no relativista a una distancia r muy grande se aproxima a esa forma. El campo eléctrico varía inversamente con r y es proporcional a la aceleración de la carga, proyectada en el "plano de visión" y esta aceleración no es la aceleración de hoy, sino la aceleración que tenía en un tiempo anterior, siendo el atraso un tiempo r/c. En lo que queda de este capítulo vamos a discutir esta ley de manera que la podamos entender mejor físicamente, porque la vamos a usar para entender todos los fenómenos de la propagación de luz y radio, como reflexión, refracción, interferencia, difracción y dispersión. Es la ley central y es todo lo que necesitamos. Todo el resto de la ecuación (28.3) fue escrito sólo para montar el escenario, de manera que pudiéramos apreciar dónde encaja (28.6) y cómo surge.

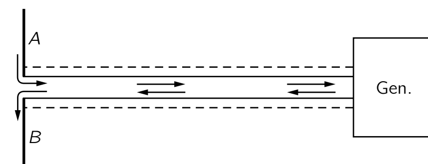


Figura 28-1. Un generador de señal de alta frecuencia impulsa cargas hacia arriba y hacia abajo en dos alambres.

Vamos a discutir más la ecuación (28.3) el próximo año. Mientras tanto la vamos a aceptar como verdadera, pero no precisamente sobre una base teórica. Podemos diseñar varios experimentos que ilustran el carácter de la ley. Para hacerlo, necesitamos una carga acelerada. Debería ser una sola carga, pero si podemos lograr que muchas cargas se muevan en conjunto todas de la misma manera, sabemos que el campo va a ser la suma de los efectos de cada una de las cargas individuales, sencillamente las sumamos. Como ejemplo, consideren dos pedazos de alambre conectados a un generador como se muestra en la figura 28-1. La idea es que el generador produce una diferencia de potencial o un campo, el cual arranca electrones de la porción A y los empuja hacia B en un cierto instante y luego, un tiempo infinitesimal más tarde, invierte el efecto y ¡arranca los electrones de B y los bombea de vuelta a A! De manera que en estos dos alambres las cargas, por así decir, están acelerando por un momento hacia arriba en el alambre A y en el B, y después en una porción brevísima de tiempo están acelerando hacia abajo en el alambre A y en el B. El que necesitemos dos alambres y un generador es simplemente que ésa es una manera de hacerlo. El resultado neto es que simplemente tenemos una carga acelerada hacia arriba y hacia abajo como si A y B fueran un solo alambre. Un alambre que es muy corto comparado con la distancia que viaja la luz en el periodo de una oscilación se llama *oscilador dipolar eléctrico.* Así, tenemos la condición que necesitamos para

aplicar nuestra ley, que nos dice que esta carga produce un campo eléctrico, de manera que necesitamos un instrumento para detectar un campo eléctrico y el instrumento que usamos es la misma cosa: ¡un par de alambres como A y B! Si un campo eléctrico se aplica a tal dispositivo, va a producir una fuerza que va a arrastrar los electrones hacia arriba en ambos alambres o hacia abajo en ambos alambres. Esta señal se detecta mediante un rectificador colocado entre A y B y un alambre diminuto y fino lleva la información a un amplificador donde es amplificada de manera que podamos oír el tono de audiofrecuencia con el cual la radiofrecuencia está modulada. Cuando esta sonda siente un campo eléctrico saldrá un ruido fuerte del amplificador y cuando no haya campo eléctrico, no habrá ruido.

Como el espacio en donde estamos midiendo las ondas contiene otros objetos, nuestro campo eléctrico va a agitar electrones en estos otros objetos, el campo eléctrico hace que estas cargas vayan hacia arriba y hacia abajo y al ir hacia arriba y hacia abajo también producen un efecto de nuestra sonda. Luego, para un experimento exitoso debemos mantener las cosas bastante juntas de manera que las influencias desde las paredes y desde nosotros mismos —las ondas reflejadas— sean relativamente pequeñas. Así resultará que los fenómenos parecen no estar de acuerdo en forma precisa y perfecta con la ecuación (28.6), pero serán lo suficientemente parecidos para que podamos apreciar la ley.

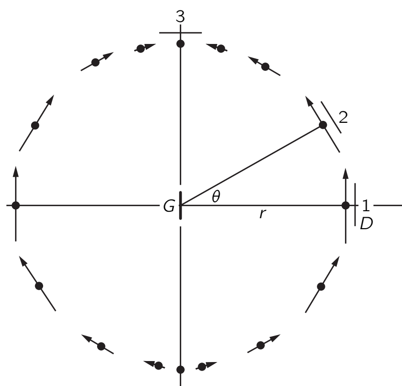


Figura 28-2. El campo eléctrico instantáneo sobre una esfera centrada en una carga localizada que oscila linealmente.

Acto seguido conectamos el generador y escuchamos la señal de audio. Encontramos un campo intenso cuando el detector D está paralelo al generador G en el punto 1, (figura 28-2). También encontramos el mismo campo en cualquier otro ángulo azimutal alrededor del eje de G, porque no tiene efectos direccionales. Por otro lado, cuando el detector está en 3, el campo es cero. Esto es correcto, porque nuestra fórmula decía que el campo debía ser la aceleración de la carga *proyectada perpendicularmente* a la visual. Por lo tanto, cuando miramos hacia G, la carga se está moviendo hacia y desde D, y no hay efecto alguno. Así que eso verifica la primera regla: que no hay efecto cuando la carga se está moviendo directamente hacia nosotros. En segundo lugar, la fórmula dice que el campo eléctrico debería ser perpendicular a r y en el plano de G y r; así que si ponemos D en 1 pero lo rotamos en 90°, no deberíamos recibir señal. Y esto es precisamente lo que encontramos, el campo eléctrico es por cierto vertical y no horizontal. Cuando movemos D, hasta un cierto ángulo intermedio, vemos que la señal más fuerte aparece cuando está orientado como se muestra, porque aunque G es vertical no produce un campo que sea simplemente paralelo a sí mismo —lo que cuenta es la *proyección de la aceleración perpendicular a la visual*—. La señal es más débil en 2 que en 1 debido al efecto de proyección.

28-4 Interferencia

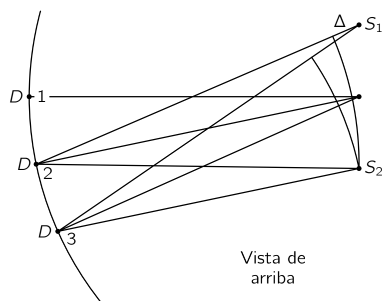


Figura 28-3. Ilustración de la interferencia de fuentes.

A continuación podemos probar lo que pasa cuando tenemos dos fuentes una al lado de la otra, separadas por varias fases (figura 28-3). La ley es que las dos fuentes deberían sumar sus efectos en el punto 1 cuando ambas fuentes están conectadas al mismo generador y ambas se mueven hacia arriba y hacia abajo de la misma forma, de manera que el campo eléctrico total es la suma de las dos y es el doble de intenso de lo que era antes.

Ahora viene una posibilidad interesante. Supongan que hacemos que ambas cargas S_1 y S_2 aceleran hacia arriba y hacia abajo pero atrasemos el tiempo de S_2 de manera que se desfasen en 180°. Entonces el campo producido por S_1 estará en una dirección y el campo producido por S_2 estará en la dirección opuesta en cualquier instante y por lo tanto *no* deberíamos obtener efecto alguno en el punto 1. La fase de la oscilación se puede ajustar muy bien mediante un conducto que lleva la señal a S_2; cambiando el largo de este conducto cambiamos el tiempo que demora la señal en llegar a S_2 y así cambiamos la fase de esa oscilación. Ajustando este largo, podemos en verdad encontrar un lugar donde ya no queda más señal ¡a pesar de que tanto S_1 como S_2 se están moviendo! El hecho que los dos se estén moviendo se puede verificar, porque si sacamos una

vemos el movimiento de la otra. De manera que las dos juntas pueden producir *cero* si todo se ajusta correctamente.

Ahora bien, es muy interesante demostrar que la suma de dos campos es en realidad una suma *vectorial*. Ya lo hemos verificado para el movimiento hacia arriba y hacia abajo, pero probemos dos direcciones no paralelas. Primero, volvemos S_1 y S_2 a la misma fase; o sea se están moviendo juntos nuevamente. Pero ahora giramos S_1 en un ángulo de 90° como se muestra en la figura 28-4. Ahora deberíamos tener en el punto 1 la suma de dos efectos, uno de los cuales es vertical y el otro horizontal. El campo eléctrico es la suma vectorial de estas dos señales en fase —ambos son intensos al mismo tiempo y pasan por cero juntos; el campo total debería ser una señal R a 45°—. Si giramos D para obtener el máximo de ruido, debería estar cerca de 45° y no vertical. Y si lo giramos en ángulo recto respecto a esa dirección, deberíamos obtener cero, que es fácil de medir. ¡Observamos precisamente tal comportamiento!

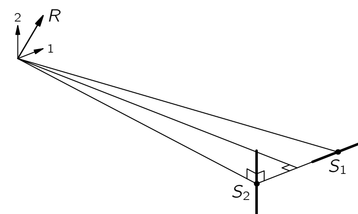


Figura 28-4. Ilustración del carácter vectorial de la combinación de fuentes.

¿Qué pasa ahora con el atraso? ¿Cómo podemos demostrar que la señal está atrasada? Podríamos, con una gran cantidad de equipo, medir el tiempo que tarda en llegar, pero hay otra manera mucho más simple. Refiriéndonos de nuevo a la figura 28-3, supongan que S_1 y S_2 estén en fase. Ambas se están sacudiendo juntas y producen campos eléctricos iguales en el punto 1. Pero supongan que vamos a un cierto lugar 2 que está más cerca de S_2 y más lejos de S_1. Entonces, de acuerdo con el principio de que la aceleración debería estar atrasada en una cantidad igual a r/c, si los atrasos no son iguales las señales ya no están más en fase. Así debería ser posible encontrar una posición en la cual las distancias a D desde S_1 y S_2 difieren en una cantidad Δ, de tal manera que no haya señal neta. Esto es, la distancia Δ debe ser la distancia que avanza la luz en media oscilación del generador. Podemos ir más allá aún y encontrar un punto donde la diferencia es mayor en un ciclo completo, o sea, la señal de la primera antena llega al punto 3 con un atraso en el tiempo que es mayor que el de la segunda antena que es justo el tiempo que demora la corriente eléctrica en oscilar una vez y, por lo tanto, los dos campos eléctricos producidos en 3 están en fase de nuevo. En el punto 3 la señal es intensa nuevamente.

Esto completa nuestra discusión de la verificación experimental de algunos aspectos importantes de la ecuación (28.6). Por cierto que no hemos verificado la variación de la intensidad del campo eléctrico con $1/r$ o el hecho de que existe un campo magnético que va junto con el campo eléctrico. Para hacerlo se necesitaría de técnicas extremadamente refinadas y difícilmente ayudaría a nuestra comprensión de este punto. En todo caso hemos verificado los aspectos que son de mayor importancia para nuestras aplicaciones posteriores y vamos a volver a estudiar algunas propiedades de las ondas electromagnéticas el próximo año.

29

Interferencia

29-1 Ondas electromagnéticas

En este capítulo discutiremos el tema del capítulo anterior en forma más matemática. Hemos demostrado cualitativamente que hay máximos y mínimos en el campo de radiación de dos fuentes, y ahora nuestro problema es describir el campo con detalles matemáticos, no sólo cualitativamente.

Ya hemos analizado físicamente el significado de la fórmula (28.6) en forma bastante satisfactoria, pero hay unos cuantos puntos a elaborar respecto a ella matemáticamente. En primer lugar, si una carga está acelerando de arriba hacia abajo a lo largo de una línea en un movimiento de amplitud muy pequeña, el campo a un cierto ángulo θ desde el eje del movimiento está en una dirección perpendicular a la visual, y en un plano que contiene tanto a la aceleración como a la visual (figura 29-1). Si la distancia se llama r, entonces en el tiempo t el campo eléctrico tiene el módulo:

$$E(t) = \frac{-qa(t - r/c)\operatorname{sen}\theta}{4\pi\epsilon_0 c^2 r}, \tag{29.1}$$

donde $a(t - r/c)$ es la aceleración en el tiempo $(t - r/c)$ llamada *aceleración retardada.*

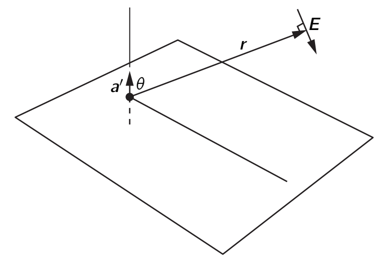


Figura 29-1. El campo eléctrico $\boldsymbol{E}$ debido a una carga positiva cuya aceleración retardada es $\boldsymbol{a}'$.

Sería interesante dibujar una imagen del campo en diferentes condiciones. Lo interesante, por supuesto, es el factor $a(t-r/c)$ y para entenderlo podemos tomar el caso más simple, $\theta = 90°$, y trazar el campo gráficamente. Lo que pensamos antes es que estamos en un lugar dado y nos preguntamos cómo varía el campo ahí con el tiempo. Pero en vez de eso vamos a ver ahora cómo es el campo en diferentes posiciones en el espacio en un instante dado. Por lo tanto, lo que queremos es una "instantánea" que nos muestre lo que es el campo en lugares diferentes. Depende por cierto de la aceleración de la carga. Supongan que la carga al principio tenía algún movimiento particular, estaba inicialmente en reposo y súbitamente aceleró de alguna manera como se muestra en la figura 29-2 y luego se detuvo. Luego, un poquito después, medimos el campo en un lugar diferente. Entonces podemos asegurar que el campo va a ser como se muestra en la figura 29-3. En cada punto el campo está determinado por la aceleración de la carga en un tiempo anterior, siendo el monto de la anticipación el atraso r/c. El campo en puntos más y más lejanos está determinado por la aceleración de la carga en tiempos más y más anteriores. Así, la curva en la figura 29-3 es realmente en cierto sentido, un diagrama "invertido" de la aceleración en función del tiempo, la distancia está relacionada con el tiempo mediante un factor de escala constante c que a menudo tomamos como unidad. Esto es fácil de ver considerando el comportamiento matemático de $a(t - r/c)$. Evidentemente, si agregamos un pequeño tiempo Δt obtenemos los mismos valores para $a(t - r/c)$ que habríamos obtenido si hubiéramos restado una distancia pequeña: $\Delta r = -c\Delta t$.

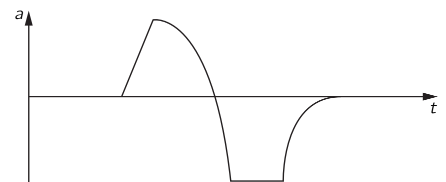


Figura 29-2. La aceleración de una cierta carga en función del tiempo.

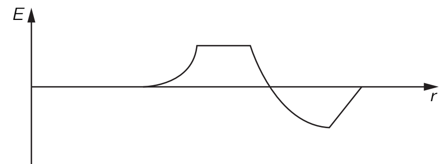


Figura 29-3. El campo eléctrico en función de la posición en un tiempo posterior. (Se ignora la variación $1/r$.)

Expresado de manera diferente, si agregamos un pequeño tiempo Δt, podemos volver $a(t-r/c)$ a su valor anterior *sumando* una pequeña distancia $\Delta r = c\Delta t$. O sea, a medida que pasa el tiempo, *el campo se mueve como una onda alejándose de la fuente.* Ésta es la razón por la cual a veces decimos que la luz se propaga como onda. Es equivalente a decir que el campo se atrasa o decir que el campo eléctrico se mueve hacia afuera mientras pasa el tiempo.

Un caso interesante es aquel en que la carga se está moviendo de arriba hacia abajo de una manera oscilatoria. En el caso que estudiamos experimentalmente en el último capítulo, el desplazamiento x en un tiempo t cualquiera era igual a

una constante x_0, la magnitud de la oscilación, por el $\cos \omega t$. Luego la aceleración es:

$$a = -\omega^2 x_0 \cos \omega t = a_0 \cos \omega t, \tag{29.2}$$

donde a_0 es la aceleración máxima $-\omega^2 x_0$. Remplazando esta fórmula en (29.1), encontramos:

$$E = -q \operatorname{sen} \theta \, \frac{a_0 \cos \omega(t - r/c)}{4\pi\epsilon_0 r c^2}. \tag{29.3}$$

Ahora, ignorando el ángulo θ y los factores constantes, veamos cómo se presenta eso en función de la posición o en función del tiempo.

29-2 Energía de radiación

Primero que nada, en cualquier momento particular o lugar particular, la intensidad del campo varía inversamente con la distancia r, como mencionamos anteriormente. Ahora debemos indicar que el contenido *energético* de una onda o los efectos energéticos que un campo eléctrico tal puede tener, son proporcionales al *cuadrado* del campo porque si, por ejemplo, tenemos algún tipo de carga o un oscilador en el campo eléctrico y si dejamos que el campo actúe sobre el oscilador lo hace moverse. Si es un oscilador lineal, la aceleración, la velocidad y el desplazamiento producido por el campo eléctrico que actúa sobre la carga son proporcionales al campo. De manera que la energía cinética que se desarrolla en la carga es proporcional al *cuadrado* del campo. Así pues, aceptaremos que la energía que un campo puede proporcionar a un sistema es proporcional de alguna manera al cuadrado del campo.

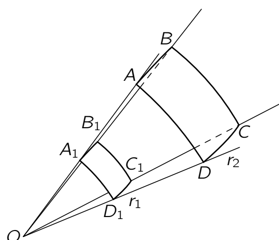


Figura 29-4. La energía que fluye dentro del cono $OABCD$ es independiente de la distancia r donde se mide.

Esto significa que la energía que la fuente puede proporcionar disminuye a medida que nos alejamos; de hecho, varía *inversamente con el cuadrado de la distancia*. Pero esto tiene una interpretación muy simple, si quisiéramos tomar de una onda toda la energía que pudiéramos en un cierto cono a una distancia r_1 (figura 29-4), y hacer lo mismo para otra distancia r_2, encontraremos que la cantidad de energía por unidad de área en cualquier lugar varía inversamente con el cuadrado de r, pero el área de la superficie interceptada por el cono varía *directamente* con el cuadrado de r. De manera que la energía que podemos sacar de la onda dentro de un ángulo cónico es la misma ¡cualquiera que sea la distancia a la que estemos! En particular, la energía total que podríamos sacar de toda la onda poniendo osciladores absorbentes en todas partes es una cierta cantidad fija. De manera que el hecho de que la amplitud de E varíe con $1/r$ es lo mismo que decir que hay un flujo de energía que no se pierde nunca, una energía que sigue y sigue extendiéndose sobre áreas efectivas más y más grandes. Vemos así que después que una carga ha oscilado ha perdido una cierta energía que nunca puede recuperar, la energía sigue alejándose más y más sin disminuir. Por lo tanto, si estamos lo suficientemente lejos de manera que nuestra aproximación básica sea suficientemente buena, la carga no puede recuperar la energía que, como decimos, ha sido radiada. Por supuesto, la energía aún existe en alguna parte y está a disposición para ser recogida por otros sistemas. Vamos a estudiar esta "pérdida" de energía en el capítulo 32.

Consideremos ahora más cuidadosamente cómo varía la onda (29.3) en función del tiempo en un lugar dado y en función de la posición en un tiempo dado. De nuevo ignoramos la variación $1/r$ y las constantes.

29-3 Ondas sinusoidaless

Primero, fijemos la posición r y observemos el campo en función del tiempo. Es oscilatorio a la frecuencia angular ω. La frecuencia angular ω puede ser definida como la *velocidad de cambio de fase con el tiempo* (radianes por segundo). Ya hemos estudiado tal cosa, así que ella debería ser muy familiar para nosotros ahora. El *periodo* es el tiempo necesario para una oscilación, un ciclo completo,

y eso también lo hemos desarrollado, pues es $2\pi/\omega$, porque ω por el periodo es un ciclo del coseno.

Ahora introducimos una nueva cantidad que se usa mucho en física. Ésta tiene que ver con la situación opuesta, en la cual fijamos t y miramos a la onda en función de la distancia r. Por supuesto, notamos que, como función de r, la onda (29.3) es también oscilatoria. Esto es, aparte del $1/r$ que estamos ignorando, vemos que E oscila cuando cambiamos de posición. Así, en analogía con ω, podemos definir una cantidad llamada *número de onda*, simbolizada por k. Éste se define como la *velocidad de cambio de fase con la distancia* (radianes por metro). O sea, a medida que nos movemos en el espacio en un tiempo fijo, la fase cambia.

Hay otra cantidad que corresponde al periodo y que podríamos llamarla periodo en el espacio, pero se llama comúnmente longitud de onda simbolizada por λ. La longitud de onda es la distancia ocupada por un ciclo completo. Es fácil ver, entonces, que la longitud de onda es $2\pi/k$, porque k por la longitud de onda sería el número de radianes en que cambia toda la cosa, o sea el producto de la velocidad de variación de los radianes por metro por el número de metros, y debemos hacer una variación de 2π para un ciclo. De manera que $k\lambda = 2\pi$ es exactamente análogo a $\omega t_0 = 2\pi$.

Ahora bien, en nuestra onda particular, hay una relación bien definida entre la frecuencia y la longitud de onda, pero las definiciones anteriores de k y ω son de hecho totalmente generales. Esto es, la longitud de onda y la frecuencia pueden no estar relacionadas de la misma manera en otras circunstancias físicas. Sin embargo, en nuestro caso la variación de la fase con la distancia se determina fácilmente, porque si llamamos $\phi = \omega(t - r/c)$ la fase y derivamos (parcialmente) respecto a la distancia r, la variación $\partial\phi/\partial r$ es:

$$\left|\frac{\partial\phi}{\partial r}\right| = k = \frac{\omega}{c}. \tag{29.4}$$

Hay muchas maneras de representar la misma cosa, como:

$$\lambda = ct_0 \tag{29.5}$$

$$\omega = ck \tag{29.6}$$

$$\lambda\nu = c \tag{29.7}$$

$$\omega\lambda = 2\pi c \tag{29.8}$$

¿Por qué es la longitud de onda igual a c por el periodo? Eso es muy fácil, por cierto, porque si nos quedamos quietos y esperamos que transcurra un periodo, las ondas, viajando a velocidad c, van a moverse una distancia ct_0 y se habrán movido por supuesto, justo en una longitud de onda.

En una situación física diferente a la de la luz, k no está relacionada necesariamente con ω de esta manera tan simple. Si llamamos x la distancia a lo largo de un eje, entonces la fórmula para una onda cosinusoidal que se mueve en una dirección x con un número de onda k y una frecuencia angular ω se escribirá en general como $\cos(\omega t - kx)$.

Ahora que hemos introducido la idea de longitud de onda podemos decir algo más acerca de los casos en los cuales (29.1) es una fórmula legítima. Recordemos que el campo está formado por varias partes, una de éstas varía inversamente con r, otra parte varía inversamente con r^2 y otras varían aún más rápido. Valdría la pena saber en qué casos la parte $1/r$ del campo es la parte más importante y las otras son relativamente pequeñas. Naturalmente, la respuesta es “si nos retiramos lo suficientemente lejos”, porque los términos que varían inversamente con el cuadrado a la larga se hacen despreciables comparados con el término $1/r$. ¿Cuánto es “suficientemente lejos”? La respuesta es cualitativamente, que los otros términos sean del orden λ/r menores que el término $1/r$. Así, mientras estemos más allá de unas cuantas longitudes de onda (29.1), es una excelente aproximación al campo. Algunas veces la región más allá de unas cuantas longitudes de onda se llama “zona de onda”.

29-4 Dos radiadores dipolares

Discutamos a continuación la matemática puesta en juego al combinar los efectos de dos osciladores para encontrar el campo resultante en un punto dado. Esto es muy fácil en los pocos casos que consideramos en el capítulo anterior. Describiremos primero los efectos cualitativamente y luego más cuantitativamente. Tomemos el caso sencillo en que los osciladores están situados con sus centros en el mismo plano horizontal que el detector y la línea de vibración es vertical.

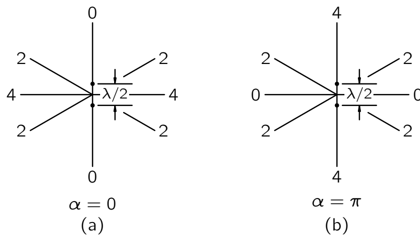


Figura 29-5. Las intensidades en varias direcciones desde dos osciladores dipolares separados media longitud de onda. Izquierda: en fase ($\alpha = 0$). Derecha: desfasados medio periodo ($\alpha = \pi$).

La figura 29-5(a) representa la vista superior de esos dos osciladores y en este ejemplo particular están separados media longitud de onda en la dirección N-S, y están oscilando con la misma fase que llamamos fase cero. Ahora nos gustaría conocer la intensidad de la radiación en diversas direcciones. Por intensidad entendemos la cantidad de energía que el campo transporta al pasar frente a nosotros por segundo, que es proporcional al cuadrado del campo promediado en el tiempo. Luego, lo que debemos mirar cuando queremos saber cuánto es el brillo de la luz, es el cuadrado del campo eléctrico, no el campo eléctrico mismo. (El campo eléctrico dice la intensidad de la fuerza que experimenta una carga estática, pero la cantidad de energía que va pasando en watt por metro cuadrado es proporcional al cuadrado del campo eléctrico. Vamos a deducir la constante de proporcionalidad en el capítulo 31). Si miramos al conjunto por el lado O, ambos osciladores contribuyen igualmente y en fase de manera que el campo eléctrico es dos veces más intenso de lo que habría sido para un oscilador único. Por lo tanto, *la intensidad es cuatro veces mayor de lo que habría sido si hubiera habido un solo oscilador*. (Los números en la figura 29-5 representan el valor que habría tenido la intensidad en este caso comparada con el que hubiera tenido si existiera allí un oscilador único de intensidad unitaria.) Ahora bien, en cualquiera de las direcciones N o S a lo largo de la línea de los osciladores, como están separados media longitud de onda, el efecto de un oscilador resulta estar fuera de fase en exactamente media oscilación respecto al otro, y por lo tanto los campos suman cero. A un cierto ángulo particular intermedio (concretamente a 30°) la intensidad es 2 y disminuye, 4, 2, 0, etc. Tenemos que aprender a encontrar estos números para otros ángulos. Se trata de sumar dos oscilaciones con fases diferentes.

Examinemos rápidamente otros casos de interés. Supongan que los osciladores estén nuevamente separados media longitud de onda, pero la fase α de uno se atrasa medio periodo con respecto a la otra en su oscilación (29-5b). En la dirección oeste la intensidad es cero ahora, porque un oscilador está "empujando" cuando el otro está "tirando". Pero en la dirección norte la señal del más cercano llega en un cierto tiempo y la del otro llega medio periodo más tarde. Pero el último estaba *originalmente* atrasado en medio periodo y por lo tanto está exactamente *a tiempo* con el primero y por lo tanto la intensidad en esta dirección es 4 unidades. La intensidad en la dirección 30° sigue siendo 2, como podemos demostrar luego.

Ahora llegamos a un caso interesante que muestra un aspecto posiblemente útil. Observemos que una de las razones por la cual las relaciones de fase de los osciladores son interesantes es que se usa esto para dirigir las transmisiones de radio. Por ejemplo, si construimos un sistema de antenas y queremos mandar una señal de radio, digamos a Hawái, ponemos las antenas como en la figura 29-5(a) y transmitimos con nuestras dos antenas en fase porque Hawái está al oeste de nosotros. Luego decidimos que mañana vamos a transmitir a Alberta, Canadá. Como está al norte, no al oeste, todo lo que tenemos que hacer es invertir la fase de una de nuestras antenas y podemos transmitir para el norte. Así podemos construir sistemas de antena con diversos arreglos. El nuestro es uno de los más simples posibles, podemos hacerlos mucho más complicados, y cambiando las fases en las diversas antenas podemos mandar los haces en muchas direcciones y enviar la mayor parte de la potencia en la dirección que queremos transmitir ¡sin mover nunca la antena! En ambos casos precedentes, sin embargo, mientras estamos transmitiendo hacia Alberta estamos gastando mucha potencia hacia la isla de Pascua y sería interesante preguntar si es posible mandarla en *una* sola

dirección. A primera vista podríamos pensar que con un par de antenas de esta naturaleza el resultado va a ser siempre simétrico. Así, pues, consideremos un caso en que sale asimétrico, para mostrar la posible variedad.

Si las antenas están separadas un cuarto de longitud de onda y si la del N está atrasada un cuarto de período en el tiempo respecto a la del S, ¿qué sucede entonces (figura 29-6)? En la dirección O obtenemos 2, como veremos luego. En la dirección S obtenemos *cero* porque la señal de S llega en un cierto tiempo, la del N llega 90° atrasada en el *tiempo*, pero ya está atrasada en 90° en su fase impuesta, luego llega con un desfase total de 180° y no hay efecto. Por otro lado, en la dirección N, la señal N llega más temprano que la señal S en 90° en el tiempo, porque está un cuarto de longitud de onda más cerca. Pero su fase está puesta de tal manera que está oscilando 90° *atrasada* en el tiempo, lo que compensa justamente la diferencia de atraso y, por lo tanto, las dos señales aparecen *juntas* en fase haciendo la intensidad del campo el doble de grande y la energía cuatro veces mayor.

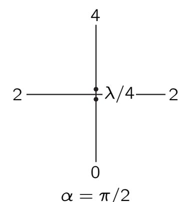


Figura 29-6. Un par de antenas dipolares que dan potencia máxima en una dirección.

Por lo tanto, usando cierta inteligencia al separar y desfasar nuestras antenas, podemos enviar toda la potencia en una dirección. Pero todavía está distribuida sobre un gran intervalo de ángulos. ¿Podemos arreglar las cosas de manera que esté enfocada aún más nítidamente en una dirección particular? Consideremos el caso de Hawái, donde estamos enviando el haz hacia el este y el oeste, pero está esparcido en un ángulo bastante grande, porque aún a 30° todavía estamos obteniendo la mitad de la intensidad —estamos malgastando potencia—. ¿Podemos mejorar eso? Tomemos la situación en que la separación es 10 longitudes de onda (figura 29-7), que es más cercanamente comparable a la situación con que experimentamos en el capítulo anterior con separaciones de varias longitudes de onda en vez de una pequeña fracción de una longitud de onda. Aquí el cuadro es bien diferente.

Si los osciladores están separados 10 longitudes de onda (tomamos el caso en fase para hacerlo más fácil), vemos que en la dirección E-O están en fase y obtenemos una gran intensidad, cuatro veces lo que habríamos obtenido si uno de ellos estuviera solo. Por otro lado, en un ángulo de separación pequeño, los tiempos de llegada difieren en 180° y la intensidad es cero. Para ser precisos, si trazamos una línea desde cada oscilador a un punto distante y la diferencia Δ en las dos distancias es $\lambda/2$, media oscilación, entonces van a estar fuera de fase. Así que este primer cero ocurre cuando eso sucede. (La figura no está dibujada a escala, es sólo un bosquejo aproximado.) Esto significa que realmente tenemos un rayo muy agudo en la dirección que queremos, porque si nos movemos sólo un poco, perderemos toda nuestra intensidad. Desgraciadamente para fines prácticos, si estuviéramos pensando en hacer un aparato de transmisión de radio y dobláramos la distancia Δ, entonces estaríamos un ciclo completo fuera de fase ¡que es lo mismo que estar *en* fase de nuevo! Así encontramos muchos máximos y mínimos sucesivos justamente como encontramos con la separación de $2\frac{1}{2}\lambda$ en el capítulo 28.

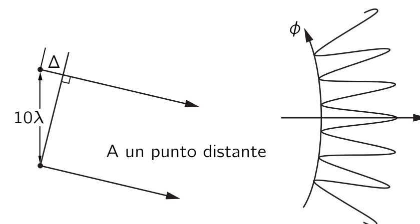


Figura 29-7. Diagrama de intensidad para dos dipolos separados 10λ.

¿Cómo podemos deshacernos de estos máximos adicionales, o "lóbulos", como se les llama? Nos podríamos deshacer de los lóbulos indeseables de una manera bastante interesante. Supongan que pusiéramos otro conjunto de antenas entre los dos que ya tenemos. O sea, las exteriores aún están separadas 10λ, pero entre ellas, digamos cada 2λ, hemos puesto otra antena y las excitamos a todas en fase. Existen ahora seis antenas y si observáramos la intensidad en la dirección E-O sería, por supuesto, mucho mayor con seis antenas que con una. El campo sería seis veces y la intensidad 36 veces mayor (el cuadrado del campo). Obtenemos 36 unidades de intensidad en esa dirección. Ahora bien, si observamos en los puntos vecinos, encontramos un cero como antes, aproximadamente, pero si avanzamos hasta donde teníamos una gran "protuberancia", obtenemos ahora una "protuberancia" mucho menor. Tratemos de ver por qué.

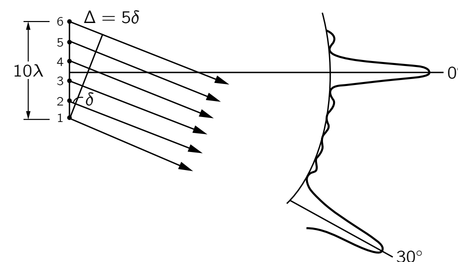


Figura 29-8. Un conjunto de seis antenas dipolares y parte de su diagrama de intensidad.

La razón es que, a pesar de que podríamos esperar una gran protuberancia cuando la distancia Δ es exactamente igual a la longitud de onda, es cierto que los dipolos 1 y 6 están entonces en fase y están cooperando en tratar de obtener alguna intensidad en esa dirección. Pero los números 3 y 4 están aproximadamente

a $\frac{1}{2}$ longitud de onda fuera de fase con 1 y 6 y, aunque 1 y 6 empujan juntos, 3 y 4 empujan juntos también, pero en la dirección opuesta. Luego hay muy poca intensidad en esa dirección —pero hay algo, ella no se compensa exactamente—. Este tipo de cosas sigue sucediendo, obtenemos protuberancias muy chicas y tenemos el haz fuerte en la dirección en que lo queríamos. Pero en este ejemplo particular, algo más va a suceder, a saber: como la distancia entre dipolos sucesivos es 2λ, es posible encontrar un ángulo donde la distancia δ entre *dipolos sucesivos* es exactamente una longitud de onda de manera que los efectos de cada uno de ellos estén de nuevo en fase. Cada uno está atrasado respecto al siguiente en 360°, de manera que todos vuelven en fase ¡y tenemos otro haz mayor en esa dirección! Es fácil evitar esto en la práctica, porque es posible poner los dipolos más cerca que a una separación de una longitud de onda. Si ponemos más antenas, más cerca que una longitud de onda, esto no puede suceder. Pero el hecho de que esto *pueda* suceder para ciertos ángulos, si la separación es mayor que una longitud de onda, es un fenómeno muy interesante y útil en otras aplicaciones —no en la transmisión de radio, sino en *redes de difracción*—.

29-5 La matemática de la interferencia

Ahora hemos terminado nuestro análisis del fenómeno de radiador dipolar cualitativamente y debemos aprender cómo analizarlos cuantitativamente. Para encontrar el efecto de dos fuentes a un cierto ángulo particular en el caso más general, donde los dos osciladores tienen un desfasaje intrínseco α y las intensidades A_1 y A_2 no son iguales, encontramos que tenemos que sumar dos cosenos que tienen la misma frecuencia pero con fases diferentes. Es muy fácil encontrar esta diferencia de fase, está constituida por el atraso debido a la diferencia de distancias y la fase intrínseca impuesta a la oscilación. Matemáticamente debemos encontrar la suma R de dos ondas: $R = A_1 \cos(\omega t + \phi_1) + A_2 \cos(\omega t + \phi_2)$ ¿Cómo hacerlo?

Es realmente muy fácil y suponemos que ya sabemos cómo hacerlo. Sin embargo, vamos a delinear el procedimiento con cierto detalle. Primero podemos, si somos hábiles en matemática y sabemos lo suficiente de cosenos y senos, desarrollarlo simplemente. El caso más simple es aquel en que A_1 y A_2 son iguales, digamos ambos iguales a A. En estas circunstancias, por ejemplo (podríamos llamarle métodos trigonométricos de resolución del problemas), tenemos:

$$R = A[\cos(\omega t + \phi_1) + \cos(\omega t + \phi_2)]. \tag{29.9}$$

Alguna vez, en nuestra clase de trigonometría, habremos aprendido la regla que,

$$\cos A + \cos B = 2\cos\tfrac{1}{2}(A + B)\cos\tfrac{1}{2}(A - B). \tag{29.10}$$

Si sabemos eso, podemos inmediatamente escribir R en la forma

$$R = 2A\cos\tfrac{1}{2}(\phi_1 - \phi_2)\cos(\omega t + \tfrac{1}{2}\phi_1 + \tfrac{1}{2}\phi_2). \tag{29.11}$$

Así, tenemos una onda oscilatoria con una nueva fase y una nueva amplitud. En general, el resultado *será* una onda oscilatoria con una nueva amplitud A_R, que podríamos llamar amplitud resultante, y que oscila con la misma frecuencia pero con un desfase ϕ_R, llamado fase resultante. En vista de esto, nuestro caso particular tiene el siguiente resultado, que la amplitud resultante es:

$$A_R = 2A\cos\tfrac{1}{2}(\phi_1 - \phi_2), \tag{29.12}$$

y la fase resultante es el promedio de las dos fases y hemos resuelto completamente nuestro problema.

Supongan ahora que no podemos recordar que la suma de dos cosenos es el doble del coseno de la semisuma por el coseno de la semidiferencia. Entonces podemos usar otro método de análisis que es más geométrico. Cualquier función coseno de ωt se puede considerar como la proyección horizontal de un *vector de rotación*. Supongan que existiera un vector $\boldsymbol{A}_1$ de largo A_1 rotando en el tiempo

de manera que el ángulo con el eje horizontal es $\omega t + \phi_1$. (Vamos a dejar de lado el ωt por un momento más y veremos que ello no tiene importancia.) Supongan que tomamos una instantánea en el tiempo $t = 0$, aunque de hecho la fotografía está rotando con velocidad angular ω (figura 29-9). La proyección de $\boldsymbol{A}_1$ según el eje horizontal es precisamente $A_1 \cos(\omega t + \phi_1)$. Ahora bien, para $t = 0$ la segunda onda podría estar representada por otro vector $\boldsymbol{A}_2$ de largo A_2 y con un ángulo ϕ_2, y que también rota. Ambos están rotando con la misma velocidad angular ω y, por lo tanto, las posiciones *relativas* de ambos son fijas. El sistema gira como un cuerpo rígido. La proyección horizontal de $\boldsymbol{A}_2$ es $A_2 \cos(\omega t + \phi_2)$. Pero sabemos por la teoría de vectores que si sumamos los dos vectores de la manera acostumbrada mediante la regla del paralelogramo y dibujamos el vector resultante $\boldsymbol{A}_R$, la componente x de la resultante es la suma de las componentes x de los otros dos vectores. Eso resuelve nuestro problema. Es fácil verificar que esto da el resultado correcto para el caso especial que consideramos más arriba donde $A_1 = A_2 = A$. En este caso vemos en la figura 29-9 que $\boldsymbol{A}_R$ está a medio camino entre $\boldsymbol{A}_1$ y $\boldsymbol{A}_2$ y forma un ángulo $\frac{1}{2}(\phi_2 - \phi_1)$ con cada uno. Por lo tanto, vemos que $A_R = 2A \cos \frac{1}{2}(\phi_2 - \phi_1)$ como antes. También, según vemos en el triángulo, la fase de $\boldsymbol{A}_R$ mientras gira, es el ángulo promedio $\boldsymbol{A}_1$ y $\boldsymbol{A}_2$ cuando las dos amplitudes son iguales. Sin duda también podemos resolver el caso en que las amplitudes no son iguales con la misma facilidad. Podemos llamar a lo anterior la manera *geométrica* de resolver el problema.

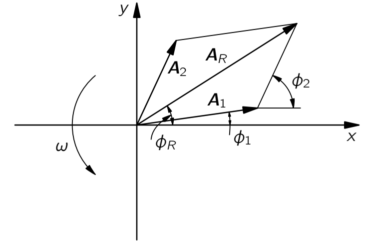


Figura 29-9. Método geométrico para combinar dos ondas cosinusoidales. Se supone que todo el diagrama está girando en sentido antihorario con frecuencia angular ω.

Hay todavía otra manera de resolver el problema, la manera *analítica*. O sea, en vez de tener que hacer realmente un dibujo como el de la figura 29-9, podemos escribir algo que diga lo mismo que el dibujo: en vez de dibujar los vectores escribimos un *número complejo* para representar cada uno de los vectores. La parte real de los números complejos son las cantidades físicas verdaderas. Luego, en nuestro caso particular, las ondas podrían escribirse de esta manera: $A_1 e^{i(\omega t + \phi_1)}$, [la parte real de esto es $A_1 \cos(\omega t + \phi_1)$] y $A_2 e^{i(\omega t + \phi_2)}$. Ahora podemos sumar los dos:

$$R = A_1 e^{i(\omega t + \phi_1)} + A_2 e^{i(\omega t + \phi_2)} = (A_1 e^{i\phi_1} + A_2 e^{i\phi_2}) e^{i\omega t} \tag{29.13}$$

o

$$\hat{R} = A_1 e^{i\phi_1} + A_2 e^{i\phi_2} = A_R e^{i\phi_R}. \tag{29.14}$$

Esto resuelve el problema que queríamos resolver, porque representa el resultado como un número complejo de módulo A_R y fase ϕ_R.

Para ver cómo funciona este método encontremos la amplitud A_R que es el "largo" de $\hat{R}$. Para obtener el "largo" de una cantidad compleja, siempre multiplicamos la cantidad por su complejo conjugado que da el largo al cuadrado. El complejo conjugado es la misma expresión, pero con el signo de i invertido. Tenemos así,

$$A_R^2 = (A_1 e^{i\phi_1} + A_2 e^{i\phi_2})(A_1 e^{-i\phi_1} + A_2 e^{-i\phi_2}). \tag{29.15}$$

Al hacer la multiplicación obtenemos $A_1^2 + A_2^2$ (aquí se cancelan las e) y para los términos cruzados tenemos:

$$A_1 A_2 (e^{i(\phi_1 - \phi_2)} + e^{i(\phi_2 - \phi_1)}).$$

Ahora bien,

$$e^{i\theta} + e^{-i\theta} = \cos\theta + i \operatorname{sen}\theta + \cos\theta - i \operatorname{sen}\theta.$$

O sea $e^{i\theta} + e^{i\theta} = 2\cos\theta$. Nuestro resultado final es, por lo tanto,

$$A_R^2 = A_1^2 + A_2^2 + 2A_1 A_2 \cos(\phi_2 - \phi_1). \tag{29.16}$$

Como vemos, esto está de acuerdo con el largo de $\boldsymbol{A}_R$ en la figura 29-9 usando las reglas de trigonometría.

Luego, la suma de los dos efectos tiene la intensidad de A_1^2 que obtendríamos con uno de ellos solo más la intensidad A_2^2 que obtendríamos con el otro solo, más una corrección. Esta corrección la llamamos *efecto de interferencia*. Es realmente

sólo la diferencia entre lo que obtenemos por simple suma de las intensidades y lo que realmente sucede. Lo llamamos interferencia, sea positiva o negativa. (Interferencia en lenguaje común usualmente sugiere oposición o impedimento, ¡pero en la física a menudo no usamos el lenguaje de la manera en que fue diseñado originalmente!) Si el término de interferencia es positivo llamamos ese caso interferencia *constructiva* ¡por muy horrible que pueda sonar a toda persona que no sea un físico! El caso opuesto se llama interferencia *destructiva*.

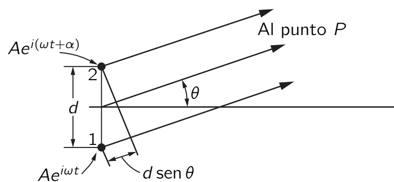


Figura 29-10. Dos osciladores de igual amplitud desfasados en α.

Ahora veamos cómo aplicar nuestra fórmula general (29.16) al caso de dos osciladores en las situaciones especiales que hemos discutido cualitativamente. Para aplicar esta fórmula general, sólo es necesario encontrar qué desfase $\phi_1 - \phi_2$ existe entre las señales que llegan a un punto dado. (Depende del desfase solamente, no de la fase misma). Así que consideremos el caso en que dos osciladores de la misma amplitud están separados por una distancia d y tienen un desfase intrínseco α. (Cuando uno tiene fase cero, la fase del otro es α.) Entonces preguntamos cuál será la intensidad en dirección acimutal θ respecto a la línea E-O. Noten que ésta *no* es la misma θ que aparece en (29.1). Estamos en duda de usar un símbolo no convencional como $\not{U}$ o el símbolo convencional θ (figura 29-10). La relación entre las fases se encuentra observando que la diferencia de distancias entre P y los dos osciladores es $d\,\mathrm{sen}\,\theta$, de manera que la contribución al desfase debida a esto es el número de longitudes de onda que hay en $d\,\mathrm{sen}\,\theta$ multiplicado por 2π. (Aquellos que son más refinados podrían querer multiplicar el número de onda k, que es la velocidad de cambio de la fase con la distancia por $d\,\mathrm{sen}\,\theta$, es exactamente lo mismo.) Luego, el desfase debido a la diferencia de distancias es $2\pi d\,\mathrm{sen}\,\theta/\lambda$, pero debido al desfase temporal de los osciladores hay una fase adicional α. Luego el desfase a la llegada sería:

$$\phi_2 - \phi_1 = \alpha + 2\pi d\,\mathrm{sen}\,\theta/\lambda. \tag{29.17}$$

Esto contiene todos los casos. Así, todo lo que tenemos que hacer es remplazar esta expresión en (29.16) para el caso $A_1 = A_2$ y podemos calcular todos los diversos resultados para dos antenas de igual intensidad.

Veamos ahora qué sucede en nuestros diversos casos. La razón que sepamos, por ejemplo, que la intensidad es 2 para 30° en la figura 29-5 es la siguiente: los dos osciladores están separados $\frac{1}{2}\lambda$, de manera que a 30°, $d\,\mathrm{sen}\,\theta = \lambda/4$. Luego $\phi_2 - \phi_1 = 2\pi\lambda/4\lambda = \pi/2$ y, por lo tanto, el término de interferencia es cero. (Estamos sumando dos vectores a 90°.) El resultado es la hipotenusa de un triángulo rectángulo de 45° que es $\sqrt{2}$ por la unidad de amplitud; elevándolo al cuadrado, obtenemos el doble de la intensidad de un solo oscilador. Todos los otros casos se pueden desarrollar de la misma manera.

30

Difracción

30-1 La amplitud resultante debida a n osciladores iguales

Este capítulo es una continuación directa del anterior, aunque se haya cambiado el nombre de *Interferencia* a *Difracción*. Nunca nadie ha sido capaz de definir la diferencia entre la interferencia y difraccións satisfactoriamente. Es sólo un problema de uso y no hay ninguna diferencia física específica importante entre ellas. Lo mejor que podemos hacer, hablando a la ligera, es decir que cuando hay sólo unas pocas fuentes que interfieren, digamos dos, entonces el resultado se llama comúnmente interferencia. Pero si hay un gran número, parece que la palabra difracción se usa más a menudo. Por lo cual, no nos vamos a preocupar si es interferencia o difracción, sino que continuamos directamente a partir de donde dejamos el tema en la mitad del último capítulo.

Entonces discutiremos ahora la situación cuando hay n osciladores igualmente espaciados, todos de la misma magnitud pero diferentes unos de otros en la fase, ya sea porque ellos están activados dando como resultado una fase diferente o porque los estamos mirando desde un ángulo tal que hay diferencia de tiempo de atraso. Por una razón u otra tenemos que sumar algo como esto:

$$R = A[\cos \omega t + \cos(\omega t + \phi) + \cos(\omega t + 2\phi) + \cdots + \cos(\omega t + (n-1)\phi)], \quad (30.1)$$

donde ϕ es el desfase entre un oscilador y el siguiente, como se ve desde una dirección particular. Específicamente, $\phi = \alpha + 2\pi d \operatorname{sen}\theta/\lambda$. Ahora debemos sumar todos los términos. Haremos esto geométricamente. El primero es de largo A y tiene fase cero. El siguiente también es de largo A y tiene una fase igual a ϕ. El siguiente nuevamente es de largo A y tiene una fase igual a 2ϕ y así sucesivamente. Por consiguiente, nos estamos moviendo evidentemente alrededor de un polígono comprendido por n lados y n equiángulos (figura 30-1).

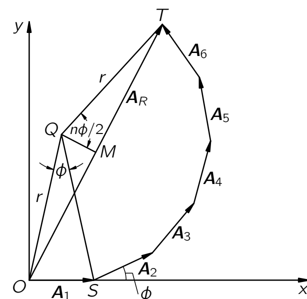


Figura 30-1. Amplitud resultante de $n = 6$ fuentes igualmente espaciadas con desfases sucesivos netos ϕ.

Ahora bien, todos los vértices están, por supuesto, sobre una circunferencia y podemos encontrar la amplitud resultante más fácilmente si encontramos el radio de esa circunferencia. Supongan que Q es el centro de la circunferencia. Entonces sabemos que el ángulo OQS es justamente un ángulo de fase ϕ. (Esto es porque el radio QS tiene la misma relación geométrica con $\boldsymbol{A}_2$ que QO con $\boldsymbol{A}_1$, de manera que forman un ángulo ϕ entre sí.) Luego el radio r debe ser tal que $A = 2r \operatorname{sen}\phi/2$, lo que da r. Pero el ángulo grande OQT es igual a $n\phi$ y encontramos así que $A_R = 2r \operatorname{sen} n\phi/2$. Combinando estos dos resultados para eliminar r, obtenemos:

$$A_R = A\,\frac{\operatorname{sen} n\phi/2}{\operatorname{sen}\phi/2}. \quad (30.2)$$

La intensidad resultante es así:

$$I = I_0\,\frac{\operatorname{sen}^2 n\phi/2}{\operatorname{sen}^2\phi/2}. \quad (30.3)$$

Analicemos ahora esta expresión y estudiemos algunas de sus consecuencias. En primer lugar la podemos verificar para $n = 1$. Ésta se verifica como $I = I_0$. Después la verificamos para $n = 2$ y se escribe $\operatorname{sen}\phi = 2\operatorname{sen}\phi/2\cos\phi/2$, encontramos que $A_R = 2A\cos\phi/2$, lo que está de acuerdo con (29.12).

Ahora bien, la idea que nos llevó a considerar la suma de varias fuentes fue que podríamos obtener una intensidad mucho mayor en una dirección que en cualquier otra, que los máximos cercanos que habrían estado presentes si hubiera

habido sólo dos fuentes, habrían disminuido en intensidad. Para ver este efecto, dibujamos la curva que proviene de (30.3), tomando n enormemente grande y trazando la región cerca de $\phi = 0$. En primer lugar, si ϕ es exactamente cero tenemos $0/0$, pero si ϕ es infinitesimal el cociente entre los dos senos al cuadrado es simplemente n^2, ya que el seno y el ángulo son aproximadamente iguales. Luego, la intensidad del máximo de la curva es igual a n^2 por la intensidad de un oscilador. Eso es fácil de ver, porque si están todos en fase, los pequeños vectores no forman ángulo relativo y todos los n se suman de manera que la amplitud es n veces mayor y la intensidad n^2 veces mayor.

A medida que la fase ϕ aumenta, el cociente entre los dos senos empieza a disminuir y la primera vez que llega a cero es cuando $n\phi/2 = \pi$, porque $\operatorname{sen}\pi = 0$. En otras palabras, $\phi = 2\pi/n$ corresponde al primer mínimo de la curva (figura 30-2). En términos de lo que sucede con las flechas en la figura 30-1, el primer mínimo sucede cuando todas las flechas vuelven al punto de partida, esto significa que el ángulo total acumulado por todas las flechas, el desfase total entre el primer oscilador y el último, debe ser 2π para completar la circunferencia.

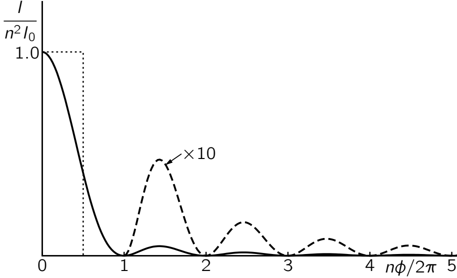


Figura 30-2. La intensidad en función del ángulo de fase para un gran número de osciladores de igual intensidad.

Ahora vamos al próximo máximo y queremos ver si es realmente mucho más chico que el primero, como habíamos esperado. No vamos a ir precisamente a la posición máxima porque tanto el numerador como el denominador de (30.3) son variables, pero $\operatorname{sen}\phi/2$ varía bastante lentamente comparado con $\operatorname{sen} n\phi/2$ cuando n es grande, de manera que cuando $\operatorname{sen} n\phi/2 = 1$, estamos muy cerca del máximo. El siguiente máximo de $\operatorname{sen}^2 n\phi/2$ se produce cuando $n\phi/2 = 3\pi/2$ o bien $\phi = 3\pi/n$. Esto corresponde a que las flechas hayan recorrido la circunferencia una vez y media. Al poner $\phi = 3\pi/n$ en la fórmula para encontrar el tamaño del máximo, encontramos que $\operatorname{sen}^2 3\pi/2 = 1$ en el numerador (por eso elegimos este ángulo) y en el denominador tenemos $\operatorname{sen}^2 3\pi/2n$. Ahora bien, si n es suficientemente grande, este ángulo es muy chico y el seno es igual al ángulo; luego, para todo fin práctico podemos poner $\operatorname{sen} 3\pi/2n = 3\pi/2n$. Encontramos así que la intensidad de este máximo es $I = I_0(4n^2/9\pi^2)$. Pero $n^2 I_0$ era la intensidad máxima y tenemos así $4/9\pi^2$ veces la intensidad máxima, lo que es cerca de 0.045, ¡menos de 5 % de la intensidad máxima! Por supuesto que hay intensidades decrecientes más allá. Por consiguiente, tenemos un máximo central muy agudo con máximos subsidiarios muy débiles a los lados.

Es posible demostrar que el área de toda la curva, incluyendo todas las pequeñas protuberancias, es igual a $2\pi n I_0$ o el doble del área del rectángulo con línea punteada en la figura 30-2.

Consideremos ahora, además, cómo podemos aplicar la ecuación (30.3) en diferentes cosas y tratemos de entender lo que sucede. Consideremos que nuestras fuentes están todas en una línea como está dibujado en la figura 30-3. Hay n de ellas, todas separadas una distancia d y vamos a suponer que el desfase intrínseco, entre una y la siguiente, es α. Luego, si estamos observando en una dirección θ de la normal, hay una fase adicional $2\pi d \operatorname{sen}\theta/\lambda$ debido al tiempo de atraso entre cada dos sucesivas, de lo que hablamos antes. Así,

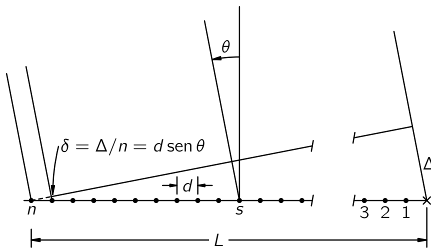


Figura 30-3. Un dispositivo lineal de n osciladores iguales activados obteniendo fases $\alpha_s = s\alpha$.

$$\begin{aligned} \phi &= \alpha + 2\pi d \operatorname{sen}\theta/\lambda \\ &= \alpha + kd \operatorname{sen}\theta. \end{aligned} \tag{30.4}$$

Primero, tomaremos el caso $\alpha = 0$. Esto es, todos los osciladores están en fase y queremos saber cuál es la intensidad en función del ángulo θ. Para averiguarlo sólo debemos poner $\phi = kd \operatorname{sen}\theta$ en la fórmula (30.3) y ver qué sucede. En primer lugar, hay un máximo cuando $\phi = 0$. Eso significa que cuando todos los osciladores están en fase hay una intensidad grande en la dirección $\theta = 0$. Por otro lado, una pregunta interesante es ¿dónde está el primer mínimo? Se produce cuando $\phi = 2\pi/n$. En otras palabras, cuando $2\pi d \operatorname{sen}\theta/\lambda = 2\pi/n$, obtenemos el primer mínimo de la curva. Si nos deshacemos de los 2π, de manera que podamos apreciar un poco mejor, esto dice que:

$$nd \operatorname{sen}\theta = \lambda. \tag{30.5}$$

Tratemos ahora de comprender físicamente por qué obtenemos un mínimo en esa posición. nd es el largo total de L. Refiriéndonos a la figura 30-3, vemos que $nd \operatorname{sen}\theta = L \operatorname{sen}\theta = \Delta$. Lo que dice (30.5) es que cuando Δ es igual a *una longitud de onda*, obtenemos un mínimo. ¿Por qué obtenemos un mínimo cuando $\Delta = \lambda$? Porque las contribuciones de los diversos osciladores están distribuidas entonces uniformemente en fase de 0° a 360°. Las flechas (figura 30-1) se desplazan alrededor de un círculo completo —estamos sumando vectores iguales en todas direcciones y esa suma es cero—. De manera que cuando tenemos un ángulo tal que $\Delta = \lambda$, obtenemos un mínimo. Ése es el primer mínimo.

Hay otro aspecto importante de la fórmula (30.3) que es que si el ángulo ϕ es aumentado en cualquier múltiplo de 2π, no introduce ninguna diferencia a la fórmula. Por consiguiente, obtenemos otros máximos grandes para $\phi = 2\pi, 4\pi, 6\pi$, etc. Cerca de cada uno de estos máximos grandes el diagrama de la figura 30-2 se repite. Nos podemos preguntar ¿cuál es la circunstancia geométrica que lleva a estos otros máximos grandes? La condición es que $\phi = 2\pi m$, donde m es cualquier entero. O sea, $2\pi d \operatorname{sen}\theta/\lambda = 2\pi m$. Dividiendo entre 2π, vemos que:

$$d \operatorname{sen}\theta = m\lambda. \tag{30.6}$$

Ésta se parece a la fórmula (30.5). No, esa fórmula era $nd \operatorname{sen}\theta = \lambda$. La diferencia es que aquí tenemos que observar a las *fuentes individuales* y cuando decimos $d \operatorname{sen}\theta = m\lambda$, eso significa que tenemos un ángulo θ tal que $\delta = m\lambda$. En otras palabras, cada fuente está ahora contribuyendo en una cierta cantidad y las sucesivas están desfasadas en un múltiplo entero de 360° y, por lo tanto, están contribuyendo *en fase*, porque estar desfasado 360° es lo mismo que estar en fase. Por consiguiente, todos contribuyen en fase y producen un máximo tan bueno como el de $m = 0$ que discutimos antes. En las protuberancias subsidiarias, toda la forma del diagrama es justamente igual a la cercana a $\phi = 0$, con exactamente los mismos mínimos a cada lado, etc. Luego, tal dispositivo va a mandar rayos en varias direcciones —cada rayo teniendo un máximo central fuerte y un cierto número de “lóbulos laterales” débiles—. Uno se refiere a los diversos rayos intensos como rayos de orden cero, primer orden, etc., de acuerdo con el valor de m. m se llama el *orden* del rayo.

Llamamos la atención sobre el hecho de que si d es menor que λ, la ecuación (30.6) no puede tener solución excepto para $m = 0$, de modo que si la separación es demasiado chica hay sólo un rayo posible, el de orden cero centrado en $\theta = 0$. (Por supuesto, también hay un rayo en la dirección opuesta.) Para obtener máximos subsidiarios grandes debemos tener la separación d del conjunto mayor que una longitud de onda.

30-2 La red de difracción

En el trabajo técnico con antenas y alambres es posible arreglar que todas las fases de los pequeños osciladores o antenas sean iguales. La pregunta es: ¿cómo podemos hacer algo similar con la luz? No podemos, en el momento presente, hacer literalmente pequeñas estaciones de radio de frecuencia óptica y unirlas con alambres infinitesimales y excitarlas todas con una fase dada. Pero hay una manera muy fácil de hacer algo que equivalga a la misma cosa.

Supongan que tuviéramos muchos alambres paralelos, igualmente espaciados con una separación d y una fuente de radiofrecuencia muy lejana, prácticamente en el infinito, que está generando un campo eléctrico que llega a cada uno de los alambres con la misma fase (está tan alejada que el atraso en el tiempo es el mismo para todos los alambres). (Uno puede desarrollar casos con dispositivos curvos, pero tomemos uno plano.) Entonces el campo eléctrico externo va a mover los electrones de arriba hacia abajo en cada alambre. Esto es, el campo que viene de la fuente original va a sacudir los electrones de arriba hacia abajo y al moverse éstos representan *nuevos generadores*. Este fenómeno se llama dispersión, una onda de luz de alguna fuente puede inducir un movimiento de los electrones en un pedazo de material y estos movimientos generan sus propias ondas. Luego, todo

lo que es necesario es disponer muchos alambres igualmente espaciados, excitarlos con una fuente de radiofrecuencia alejada y tenemos la situación que queremos, sin necesidad de muchas conexiones especiales. Si la incidencia es normal, las fases van a ser iguales, y obtendremos exactamente el caso que hemos estado discutiendo. Luego, si la separación de los alambres es mayor que la longitud de onda, obtendremos una fuerte intensidad de dispersión en la dirección normal y en ciertas otras direcciones dadas por (30.6).

¡Esto también se puede hacer con la luz! En vez de alambres, usamos un pedazo plano de vidrio y hacemos ranuras en él, tales que cada una de las ranuras disperse en forma ligeramente diferente el resto del vidrio. Si luego iluminamos el vidrio, cada una de las ranuras representará una fuente y si separamos las líneas en forma muy fina, pero no más cerca de una longitud de onda (lo que es de todos modos casi técnicamente imposible), entonces esperaríamos un fenómeno milagroso, la luz no sólo va a pasar en línea recta, sino que también habrá un fuerte haz en un ángulo distinto de cero que depende de la separación de las ranuras. Tales objetos se han hecho realmente y son de uso común —se llaman *redes de difracción*—.

En una de sus formas, una red de difracción consiste simplemente en una lámina de vidrio plano transparente e incolora con rayas. Hay a menudo varios cientos de rayas por milímetro dispuestas *muy* cuidadosamente para que estén igualmente separadas. El efecto de esta red se puede observar disponiendo un proyector para que proyecte una angosta línea vertical de luz (la imagen de una rendija) en una pantalla. Cuando ponemos la red en el haz, con los rayos verticales, vemos que la línea todavía está ahí, pero además a cada lado tenemos *otra* mancha intensa de luz que está *coloreada*. Ésta, por supuesto, es la imagen de la rendija extendida en un amplio intervalo angular, porque el ángulo θ en (30.6) depende de λ y luces de diferentes colores, como sabemos, corresponden a diferentes frecuencias y por lo tanto a diferentes longitudes de onda. La longitud de onda más larga visible es el rojo y como $d \operatorname{sen} \theta = \lambda$, eso requiere un ángulo θ mayor. ¡Y efectivamente encontramos que el rojo forma un ángulo mayor desde la imagen central! Debería haber un rayo en el otro lado también y efectivamente vemos uno en la pantalla. Luego, podría haber otra solución de (30.6) cuando $m = 2$. Vemos efectivamente que hay algo vago ahí —muy débil— y hay todavía otros rayos más allá.

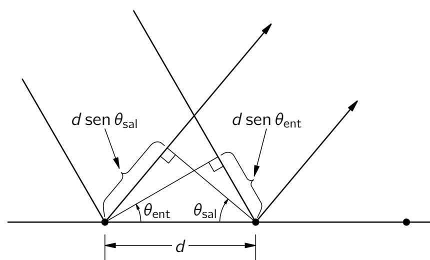


Figura 30-4. La diferencia de trayectoria para rayos dispersados por rayas adyacentes de una red es $d \operatorname{sen} \theta_{\text{sal}} - d \operatorname{sen} \theta_{\text{ent}}$.

Recién hemos dicho que todos estos rayos deberían ser de la misma intensidad, pero vemos que en realidad no lo son y en efecto ¡ni siquiera los primeros a la derecha y a la izquierda son iguales! La razón es que la red ha sido construida cuidadosamente para hacer precisamente esto. ¿Cómo? Si la red consiste en rayas muy finas, de ancho infinitesimal, separadas uniformemente, entonces las intensidades serían realmente iguales. Pero de hecho, aunque hemos tomado el caso más simple, también podríamos haber considerado un dispositivo de *pares* de antenas en el cual cada integrante del par tiene una cierta intensidad y cierta fase relativa. En este caso, es posible obtener intensidades que son diferentes en los distintos órdenes. Una red se hace a menudo con pequeños cortes de “dientes de sierra” en vez de pequeñas muescas simétricas. Disponiendo cuidadosamente los “dientes de sierra”, se puede enviar más luz a un orden particular del espectro que a otros. En una red real desearíamos tener toda la luz posible en uno de los órdenes. Esto podría parecer difícil de conseguir, pero es algo que conviene mucho realizar porque hace la red más útil.

Hasta ahora hemos considerado el caso en que las fases de las fuentes son iguales. Pero también tenemos una fórmula para ϕ cuando las fases difieren entre sí en un ángulo α. Esto requiere conectar nuestras antenas con una pequeña diferencia de fase entre sí. ¿Podemos hacer esto con la luz? Sí, lo podemos hacer muy fácilmente, para lo cual supongan que existe una fuente de luz en el infinito *a un ángulo* tal que la luz llega en un ángulo θ_{ent}, y digamos que deseamos discutir el haz dispersado que se aleja en un ángulo θ_{sal} (figura 30-4). El θ_{sal} es el mismo θ que teníamos antes, pero el θ_{ent} es sólo un medio para arreglar que la fase de cada fuente sea diferente, la luz proveniente de la fuente excitadora lejana toca primero una raya, luego otra, luego la otra y así sucesivamente con un corrimiento

de fase, de una a la otra que es como vemos $\alpha = -2\pi d \operatorname{sen}\theta_{\text{ent}}/\lambda$. Por lo tanto, tenemos la fórmula para una red en la cual la luz llega y se aleja a un ángulo:

$$\phi = 2\pi d \operatorname{sen}\theta_{\text{sal}}/\lambda - 2\pi d \operatorname{sen}\theta_{\text{ent}}/\lambda. \tag{30.7}$$

Tratemos de averiguar dónde obtenemos gran intensidad en este caso. La condición para intensidad grande es, por supuesto, que θ sea un múltiplo de 2π. Hay varios puntos interesantes que deben notarse.

Un caso de interés bastante grande es el que corresponde a $m = 0$, donde d es menor que λ; en realidad, ésta es la única solución. En este caso vemos que $\operatorname{sen}\theta_{\text{sal}} = \operatorname{sen}\theta_{\text{ent}}$, lo que puede significar θ_{sal} es el *suplemento* de θ_{ent}, por lo que la luz sale en la *misma dirección* que la luz que estaba excitando la red. Podríamos pensar que la luz "pasa de largo". No, estamos hablando de *luz diferente*. La luz que pasa de largo proviene de la fuente original, de lo que estamos hablando es de la nueva luz *que se genera por dispersión*. Resulta que la luz dispersada va en la misma dirección que la luz original; en realidad, puede interferir con ella —un aspecto que vamos a estudiar más tarde—.

Hay otra solución para este mismo caso: θ_{ent} puede ser igual a θ_{sal}. Así que no sólo vamos a obtener un rayo en la misma dirección que el rayo que llega, sino también otro en otra dirección, el cual, si lo consideramos cuidadosamente, es tal que *el ángulo de incidencia es igual al ángulo de dispersión*. Lo llamamos rayo reflejado.

Así empezamos a entender el mecanismo básico de la reflexión: la luz que llega genera movimiento de los átomos en el reflector y entonces el reflector genera una *nueva onda* y una de las soluciones para la dirección de dispersión, la *única* solución si la separación de los dispersores es pequeña comparada con una longitud de onda, es que ¡el ángulo con que sale la luz es igual al ángulo con que llega!

Enseguida, discutimos el caso especial cuando $d \to 0$. O sea, tenemos precisamente un pedazo sólido de material, por decirlo así, pero de largo finito. Además, queremos que el desfase entre un dispersor y el siguiente tienda a cero. En otras palabras, ponemos cada vez más antenas entre las anteriores de manera que cada uno de los desfases se vaya haciendo menor, pero el número de antenas crezca en tal manera que el desfase total, entre un extremo de la línea y el otro, sea constante. Veamos qué sucede con (30.3) si mantenemos constante el desfase $n\phi$ de un extremo al otro (digamos $n\phi = \Phi$), haciendo que el número tienda a infinito y el desfase ϕ de cada una tienda a cero. Pero ahora ϕ es tan pequeño que $\operatorname{sen}\phi = \phi$ y si también reconocemos $n^2 I_0$ como I_m, la intensidad máxima en el centro del rayo, encontramos:

$$I = 4I_m \operatorname{sen}^2 \tfrac{1}{2}\Phi/\Phi^2. \tag{30.8}$$

Este caso límite es lo que se muestra en la figura 30-2.

En estas circunstancias encontramos el mismo tipo general de figura que para separación finita con $d > \lambda$, todos los lóbulos laterales son prácticamente los mismos que antes, pero no hay máximos de orden superior. Si los dispersores están todos en fase, obtenemos un máximo en la dirección $\theta_{\text{sal}} = 0$ y un mínimo cuando la distancia Δ es igual a λ, como para d y n finitos. Así que hasta podemos analizar una distribución *continua* de dispersores u osciladores, usando integrales en vez de suma.

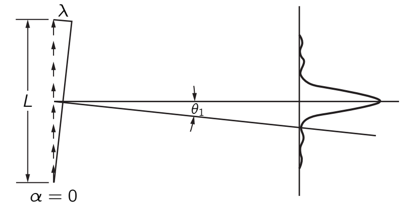


Figura 30-5. El diagrama de intensidad de una fila continua de osciladores tiene un único máximo grande y muchos "lóbulos laterales" débiles.

Como ejemplo, supongan que hubiera una larga fila de osciladores, con la carga oscilando según la dirección de la fila (véase figura 30-5). En tal dispositivo la intensidad mayor es perpendicular a la fila. Hay un poco de intensidad arriba y abajo del plano ecuatorial, pero es insignificante. Con este resultado podemos manejar una situación más complicada. Supongan que tenemos un conjunto de estas filas, cada una produciendo un rayo sólo en un plano perpendicular a la fila. Encontrar la intensidad en distintas direcciones provenientes de una serie de alambres largos, en vez de alambres infinitesimales, es el mismo problema que teníamos para alambres infinitesimales, mientras nos encontremos en el plano central perpendicular a los alambres; justamente, sumamos la contribución de

cada uno de los alambres largos. Por esto, aunque hayamos analizado en realidad sólo pequeñas antenas, podríamos igualmente bien haber usado una red con ranuras largas y delgadas. Cada una de las ranuras largas produce solamente un efecto en su propia dirección, no de arriba hacia abajo, pero están puestas una al lado de la otra horizontalmente, de modo que producen interferencia así.

Por lo tanto, podemos construir situaciones más complicadas teniendo diversas distribuciones de dispersores en filas, planos o en el espacio. Lo primero que hicimos fue considerar los dispersores en fila y hemos extendido recién el análisis a franjas; lo podemos resolver haciendo las sumas necesarias, sumando las contribuciones de los dispersores individuales. El principio es siempre el mismo.

30-3 Poder de resolución de una red

Ahora estamos en condiciones de entender numerosos fenómenos interesantes. Por ejemplo, consideren el uso de una red para separar longitudes de onda. Notamos que todo el espectro estaba separado en la pantalla, de manera que una red puede usarse como un instrumento para separar la luz en sus diversas longitudes de onda. Una de las preguntas interesantes es: suponiendo que hubiera dos fuentes de frecuencias ligeramente diferentes o de longitud de onda ligeramente distintas, ¿qué separación mínima en longitud de onda podrían tener, de manera que la red fuera incapaz de decir que realmente existían allí dos longitudes de onda diferentes? El rojo y el azul estaban separados claramente. Pero cuando una es roja y la otra es ligeramente más roja, muy próxima, ¿cuán próximas pueden estar? Esto se llama *poder de resolución* de la red, y una manera de analizar el problema es la siguiente. Supongan que para luz de un cierto color sucede que tenemos el máximo de haz difractado en un cierto ángulo. Si cambiamos la longitud de onda, la fase $2\pi d \operatorname{sen}\theta/\lambda$ es diferente, de manera que el máximo se produce por cierto a un ángulo diferente. Por esto es que el rojo y el azul están separados. ¿Qué diferencia de ángulo debe haber para que seamos capaces de verlo? Si los dos máximos están uno encima del otro, por supuesto que no los podemos ver. Si el máximo de uno está suficientemente alejado del otro, entonces vemos que hay una doble protuberancia en la distribución de la luz. Para poder discernir la doble protuberancia se usa generalmente el siguiente criterio simple llamado *criterio de Rayleigh* (figura 30-6). Éste consiste en que el primer mínimo de una protuberancia debe estar sobre el máximo de la otra. Ahora es muy fácil calcular cuál es la diferencia de longitud de onda, cuándo un mínimo está sobre el máximo de la otra. La mejor manera de hacerlo es geométricamente.

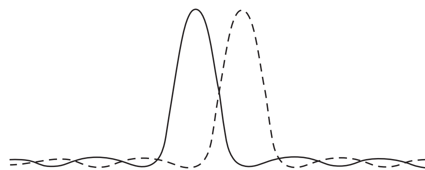


Figura 30-6. Ilustración del criterio de Rayleigh. El máximo de un diagrama cae en el primer mínimo del otro.

Para tener un máximo para la longitud de onda λ', la distancia Δ (figura 30-3) debe ser $n\lambda'$ y si estamos observando el rayo de orden m, es $mn\lambda'$. En otras palabras, $2\pi d \operatorname{sen}\theta/\lambda' = 2\pi m$; por lo tanto, $nd \operatorname{sen}\theta$, que es Δ, es $m\lambda'$ por n o $mn\lambda'$. Para el otro rayo, de longitud de onda λ, queremos tener un *mínimo* a este ángulo. Esto es, queremos que Δ sea exactamente una longitud de onda λ más que $mn\lambda$. O sea, $\Delta = mn\lambda + \lambda = mn\lambda'$. Luego si $\lambda' = \lambda + \Delta\lambda$, encontramos

$$\Delta\lambda/\lambda = 1/mn. \tag{30.9}$$

El cociente $\lambda/\Delta\lambda$ se llama *poder de resolución* de una red, vemos que es igual al número total de líneas de la red multiplicado por el orden. No es difícil probar que esta fórmula es equivalente a la fórmula que el error en la frecuencia es igual a la inversa de la diferencia de tiempo entre trayectorias extremas que se permite interferir:*

$$\Delta\nu = 1/T.$$

De hecho, ésta es la mejor manera de recordarla, porque la fórmula general funciona no sólo para redes, sino para cualquier otro instrumento, mientras que la fórmula especial (30.9) depende del hecho que estamos usando una red.

* En nuestro caso $T = \Delta/c = mn\lambda/c$, donde c es la velocidad de la luz. La frecuencia $\nu = c/\lambda$, así que $\Delta\nu = c\Delta\lambda/\lambda^2$.

30-4 La antena parabólica

Consideremos ahora otro problema en relación con el poder de resolución. Esto tiene que ver con la antena de un radio-telescopio, usado para determinar la posición de fuentes de radio en el cielo, es decir, cuán grandes son ellas en ángulo. Por supuesto, si usáramos cualquier antena común y encontráramos las señales, no sabríamos de qué dirección venían. Estamos muy interesados en saber si la fuente está en un lugar o en otro. Una manera de averiguarlo es colocar toda una serie de alambres dipolares igualmente espaciados en el paisaje australiano. Luego tomamos todos los alambres de estas antenas y los conectamos al mismo receptor de tal manera que los atrasos en las líneas de alimentación sean iguales. Así, el receptor recibe señales de todos los dipolos en fase. O sea, suma todas las ondas de cada uno de los dipolos en la misma fase. ¿Qué sucede ahora? Si la fuente está directamente encima del dispositivo, en el infinito, o aproximadamente así, entonces sus ondas de radio van a excitar a las antenas en la misma fase de manera que todas alimentan al receptor en conjunto.

Supongan ahora que la fuente de radio forme un pequeño ángulo θ con la vertical. Entonces, las diversas antenas están recibiendo señales un poco desfasadas. El receptor suma todas estas señales desfasadas, y por lo tanto no obtenemos nada si el ángulo θ es demasiado grande. ¿Qué tamaño puede tener el ángulo? *Respuesta*: obtenemos cero si el ángulo $\Delta/L = \theta$ (figura 30-3) corresponde a un desfase de $360°$, o sea si Δ es la longitud de onda λ. Esto se debe a que las contribuciones de los vectores forman en conjunto un polígono completo con resultante cero. El menor ángulo que se puede resolver mediante un dispositivo de antenas de largo L es $\theta = \lambda/L$. Noten que el diagrama de recepción de una antena como ésta es exactamente igual que la distribución de intensidad que obtendríamos si invirtiéramos el receptor y lo transformáramos en un emisor. Esto es un ejemplo de lo que se llama *principio de reciprocidad.* De hecho resulta válido en general para cualquier arreglo de antenas, ángulos, etc.; es decir, si primero resolvemos cuáles serían las intensidades relativas en diversas direcciones si el receptor fuera en su lugar un transmisor, entonces la sensibilidad direccional relativa de un receptor con las mismas conexiones externas, el mismo arreglo de antenas, es la misma que lo que sería la intensidad relativa de emisión si fuera un transmisor.

Algunas antenas de radio están hechas de otro modo. En vez de tener un montón de dipolos en una larga línea, con muchos alambres de alimentación, podemos disponerlos no en una línea, sino en una curva y poner el receptor en un cierto punto donde pueda detectar las ondas dispersadas. Esta curva está diseñada inteligentemente de manera que si las ondas de radio llegan desde arriba y los alambres dispersan formando una nueva onda, los alambres están dispuestos de tal modo que las ondas dispersadas llegan al receptor todas al mismo tiempo (figura 26-12). En otras palabras, la curva es una *parábola* y cuando la fuente esté exactamente en su eje, obtenemos una intensidad muy grande en el foco. En este caso entendemos muy claramente cuál es el poder de resolución de este instrumento. El arreglo de las antenas en una curva parabólica no es una cuestión esencial. Es sólo una manera conveniente de llevar las señales a un mismo punto sin atrasos relativos y sin alambres de alimentación. El ángulo que ese instrumento puede resolver es todavía $\theta = \lambda/L$, donde L es la separación entre la primera y la última antena. No depende de la separación de las antenas que pueden estar muy juntas o, de hecho, ser un pedazo de metal. Ahora estamos describiendo un espejo telescópico, por supuesto. ¡Encontramos el poder de resolución de un telescopio! (A veces el poder de resolución se escribe $\theta = 1.22\lambda/L$, donde L es el diámetro del telescopio. La razón que no sea exactamente λ/L es ésta: cuando calculamos que $\theta = \lambda/L$, supusimos que todas las líneas de dipolos eran iguales en intensidad, pero cuando tenemos un telescopio circular que es la manera en que corrientemente disponemos un telescopio, no llega tanta señal desde los bordes exteriores, porque no es como el caso de un cuadrado, donde obtenemos la misma intensidad a lo largo de todo un lado. Recibimos algo menos, porque estamos usando sólo una parte del telescopio ahí, podemos así apreciar que el diámetro

efectivo es un poco más corto que el diámetro verdadero, y eso es lo que nos dice el factor 1.22. En todo caso parece un poco pedante poner tanta precisión en la fórmula del poder de resolución)*.

30-5 Películas coloreadas, cristales

Lo anterior, entonces, constituyen algunos de los efectos de interferencia que se obtienen al sumar las diversas ondas. Pero hay una serie de otros ejemplos, y a pesar de que aún no entendamos el mecanismo fundamental, algún día lo entenderemos y podemos comprender incluso ahora cómo sucede la interferencia. Por ejemplo, cuando una onda de luz choca con la superficie de un material de índice n, digamos en incidencia normal, algo de la luz se refleja. No estamos ahora mismo en situación de comprender la *razón* de la reflexión, la discutiremos más tarde. Pero supongamos que sabemos que algo de luz se refleja tanto al entrar como al salir de un medio refractante. Luego, si miramos a la reflexión de una fuente luminosa en una película delgada, vemos la suma de dos ondas; si el espesor es suficientemente pequeño, estas dos ondas van a producir interferencia, ya sea constructiva o destructiva, dependiendo de los signos de las fases. Podría ser, por ejemplo, que para la luz roja obtengamos una reflexión aumentada, pero para la luz azul, que tiene una longitud de onda diferente, obtenemos quizás una interferencia de reflexión destructiva; por lo tanto, vemos una brillante reflexión roja. Si cambiamos el espesor, es decir, si miramos en otro lugar donde la película es más gruesa, aquello puede invertirse, interfiriendo el rojo y no el azul, por lo que será azul brillante, o verde, o amarilla, o lo que sea. Así que vemos *colores* cuando miramos películas delgadas y los colores cambian si miramos desde ángulos diferentes. Así, súbitamente, apreciamos otros cientos de miles de situaciones que encierran los colores que vemos en películas de aceite, burbujas de jabón, etc., en diferentes ángulos. Pero el principio es el mismo, sólo estamos sumando ondas con fases diferentes.

Como otra importante aplicación de la difracción podemos mencionar lo siguiente. Hemos usado una red y vimos la imagen difractada en la pantalla. Si hubiéramos usado luz monocromática, habría estado en un cierto lugar específico. Luego había también varias imágenes de orden superior. De las posiciones de las imágenes podíamos decir cuánto estaban separadas las líneas de la red si conociéramos la longitud de onda de la luz. Por la diferencia de intensidad de las distintas imágenes podríamos encontrar la forma de las rayas de la red, bien que la red estuviera hecha de alambres, muescas de dientes sierra, o lo que sea, *sin poder verlas*. Este principio se usa para descubrir las posiciones de los *átomos en un cristal*. La única complicación es que un cristal es tridimensional; es un arreglo tridimensional repetido de átomos. No podemos usar luz ordinaria, porque debemos usar algo cuya longitud de onda sea menor que el espacio entre los átomos o no obtenemos ningún efecto; así que debemos usar radiación de longitud de onda muy corta, es decir, rayos X. Así, haciendo incidir rayos X en un cristal y notando lo intensa que es la reflexión en los diferentes órdenes, podemos determinar el arreglo de los átomos dentro ¡sin que podamos verlos jamás con el ojo! Es en esta forma que conocemos el arreglo de los átomos en diversas sustancias, lo que nos permitió trazar aquellos dibujos en el primer capítulo, mostrando el arreglo de los átomos en la sal, etc. Volveremos más tarde a este tema para discutirlo en mayor detalle y, por lo tanto, no diremos más por ahora acerca de esta notabilísima idea.

30-6 Difracción por pantallas opacas

Ahora llegamos a una situación muy interesante. Supongan que tenemos una lámina opaca con agujeros y una luz de un lado de ella. Queremos saber cuál es

* Esto se debe en primer lugar a que el criterio de Rayleigh es una idea aproximada. Establece cuando empieza a ser muy difícil decir si la imagen estaba formada por una o dos estrellas. En realidad si se pueden hacer medidas suficientemente cuidadosas de la distribución exacta de intensidad sobre la imagen difractada, el hecho que dos fuentes forman la imagen puede ser demostrado aun cuando θ es menor que λ/L.

la intensidad del otro lado. Lo que la mayor parte de la gente dice es que la luz ilumina a través de los agujeros y produce un efecto del otro lado. Resultará que uno obtiene la respuesta correcta, con excelente aproximación, si se supone que hay fuentes distribuidas con densidad uniforme a lo largo de los agujeros abiertos y que las fases de estas fuentes son las mismas que habrían sido si el material opaco hubiera estado ausente. Por supuesto que *no* existen realmente fuentes en los agujeros; en realidad, ése es el único lugar donde *con certeza* no hay fuentes. Sin embargo, obtenemos los diagramas de difracción correctos considerando que los agujeros son los únicos lugares donde *hay* fuentes; esto es un hecho más bien peculiar. Vamos a explicar después por qué esto es cierto, pero por ahora sólo supongamos que lo es.

En la teoría de la difracción existe otro tipo de difracción que deberíamos discutir brevemente. Por lo común no se discute tan temprano en un curso elemental como éste, sólo porque las fórmulas matemáticas necesarias para sumar estos pequeños vectores son un poco complicadas. Fuera de eso, es exactamente lo mismo que hemos estado haciendo todo el tiempo. Todos los fenómenos de interferencia son los mismos, no se incluye nada mucho más avanzado, sólo las circunstancias son más complicadas y es más difícil sumar los vectores, eso es todo.

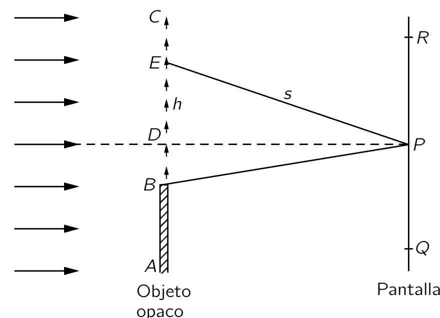


Figura 30-7. Una fuente de luz distante proyecta la sombra de un objeto opaco sobre una pantalla.

Supongan que tenemos una luz que viene del infinito proyectando la sombra de un objeto. La figura 30-7 muestra una pantalla en la cual la sombra de un objeto AB está producida por una fuente de luz muy lejana comparada con una longitud de onda. Podríamos ahora esperar que por fuera de la sombra esté todo iluminado y por dentro todo oscuro. De hecho, si graficamos la intensidad en función de la posición cerca del borde de la sombra, la intensidad aumenta y luego se sobrepasa, se bambolea y oscila de una manera muy peculiar cerca de este borde (figura 30-9). Ahora vamos a discutir la razón de esto. Si usamos el teorema que no hemos demostrado todavía, podemos remplazar el problema real por un conjunto de fuentes efectivas distribuidas uniformemente en el espacio abierto más allá del objeto.

Imaginamos un gran número de antenas muy poco distantes y deseamos la intensidad en un cierto punto P. Esto se parece justamente a lo que hemos estado haciendo. No totalmente; porque nuestra pantalla no está en el infinito. No queremos la intensidad en el infinito, sino en un punto finito. Para calcular la intensidad en un lugar particular, tenemos que sumar las contribuciones de todas las antenas. Primero, hay una antena en D, opuesta exactamente a P, si subimos un poco en ángulo, digamos una altura h; entonces hay un aumento en el atraso (también hay un cambio en la amplitud debido al cambio en la distancia, pero éste es un efecto muy pequeño si estamos lejos y es mucho menos importante que el desfase). La diferencia en la trayectoria $EP - DP$ es aproximadamente $h^2/2s$, de manera que el desfase es proporcional al *cuadrado* de cuánto nos alejamos de D, mientras que en nuestro trabajo previo era infinito y el desfase era *linealmente* proporcional a h. Cuando las fases son linealmente proporcionales, cada vector se suma formando un ángulo constante con el siguiente. Lo que necesitamos ahora es una curva que se hace sumando muchos vectores infinitesimales con la condición de que el ángulo que forman aumente, no linealmente, sino con el *cuadrado* de la longitud de la curva. Construir esa curva pone en juego una matemática ligeramente avanzada, pero siempre la podemos construir dibujando efectivamente las flechas y midiendo los ángulos. En todo caso, obtenemos la maravillosa curva (llamada espiral de Cornu) que se muestra en la figura 30-8. Ahora bien, ¿cómo usamos esta curva?

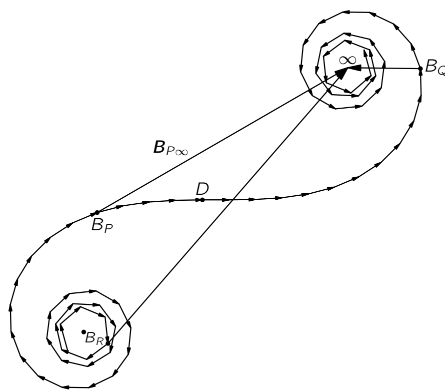


Figura 30-8. La suma de amplitudes para muchos osciladores en fase cuyos retardos en las fases varían con el cuadrado de la distancia desde el punto D de la figura anterior.

Si queremos la intensidad, digamos en el punto P, sumamos una serie de contribuciones de fases diferentes desde el punto D hacia arriba hasta infinito y desde D hacia abajo sólo hasta el punto B_P. De manera que partimos de B_P en la figura 30-8 y dibujamos una serie de flechas con ángulos siempre crecientes. Por lo tanto, la contribución total por encima del punto B_P va a lo largo de la curva en espiral. Si dejáramos de integrar en algún lugar, la amplitud total sería un vector desde B hasta ese punto, en este problema particular vamos hasta infinito, de manera que la respuesta total es el vector $\boldsymbol{B}_{P\infty}$. Ahora bien, la posición sobre

la curva que corresponde al punto B_P en el objeto depende de dónde esté ubicado P, ya que el punto D, el punto de inflexión, siempre corresponde a la posición del punto P. Así, dependiendo de dónde esté ubicado P sobre B, el punto de partida va a caer en diferentes posiciones sobre la parte izquierda inferior de la curva y el vector resultante $\boldsymbol{B}_{P\infty}$ va a tener muchos máximos y mínimos (figura 30-9).

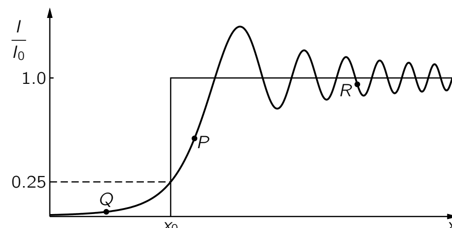


Figura 30-9. La intensidad cerca del borde de una sombra. El borde de la sombra geométrica está en x_0.

Por otro lado, si estamos en Q, al otro lado de P, estamos usando sólo un extremo de la curva espiral y no el otro extremo. En otras palabras, ni siquiera empezamos en D sino en BQ, así que en este lado obtenemos una intensidad que disminuye continuamente a medida que Q se aleja más en la sombra.

Un punto que podemos calcular inmediatamente con facilidad para mostrar que realmente lo entendemos es la intensidad exactamente opuesta al borde. La intensidad es aquí 1/4 de la luz incidente. Razón: exactamente en el borde (por consiguiente, el extremo B de la flecha está en D en la figura 30-8) tenemos la mitad de la curva que habríamos tenido si hubiéramos estado lejos en la región brillante. Si nuestro punto R está lejos dentro de la luz, vamos desde un extremo de la curva al otro, es decir, un vector unitario completo; pero si estamos en el borde de la sombra, tenemos sólo la mitad de la amplitud (1/4 de la intensidad).

En este capítulo hemos encontrado la intensidad producida en diferentes direcciones a partir de diversas distribuciones de fuentes. Como ejemplo final deduciremos una fórmula que vamos a necesitar para el próximo capítulo sobre la teoría del índice de refracción. Hasta este punto las intensidades relativas han sido suficientes para nuestros fines, pero esta vez encontraremos la fórmula completa del campo en la situación siguiente.

30-7 El campo de un plano de cargas oscilantes

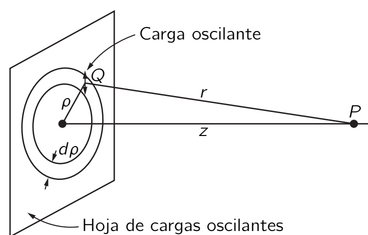


Figura 30-10. Campo de radiación de una lámina de cargas oscilantes.

Supongan que tenemos un plano lleno de fuentes, todas oscilando juntas, con su movimiento en un plano y todas con la misma amplitud y fase. ¿Cuál es el campo a una distancia finita, pero muy grande a partir del plano? (No nos podemos acercar mucho, por supuesto, porque no tenemos las fórmulas correctas para el campo cerca de las fuentes.) Si dejamos que el plano de las cargas sea el xy, queremos encontrar el campo en un punto P lejos en el eje z (figura 30-10). Suponemos que hay n cargas por unidad de área del plano y cada una de ellas tiene una carga q. Todas las cargas se mueven con movimiento armónico simple con la misma dirección, amplitud y fase. Hacemos que el movimiento de cada carga, *respecto a su propia posición intermedia*, sea $x_0 \cos \omega t$. O bien, usando la notación compleja y recordando que la parte real representa el movimiento verdadero, el movimiento puede ser descrito por $x_0 e^{i\omega t}$.

Ahora encontramos el campo en el punto P de todas las cargas, encontrando el campo ahí de cada carga q y después sumando las contribuciones de todas las cargas. Sabemos que el campo de radiación es proporcional a la aceleración de la carga que es $-\omega^2 x_0 e^{i\omega t}$ (y es la misma para cada carga). El campo eléctrico que deseamos en el punto P debido a la carga en el punto Q es proporcional a la aceleración de la carga q, pero debemos recordar que el campo en el punto P en el instante t está dado por la aceleración de la carga en el tiempo anterior $t' = t - r/c$, donde r/c es el tiempo que demoran las ondas en viajar la distancia r desde Q a P. Por lo tanto, el campo en el punto P es proporcional a

$$-\,\omega^2 x_0 e^{i\omega(t-r/c)}. \tag{30.10}$$

Usando este valor para la aceleración vista desde P en nuestra fórmula para el campo eléctrico a distancias grandes de una carga que irradia, obtenemos:

$$\begin{pmatrix}\text{Campo eléctrico en } P\\ \text{de carga en } Q\end{pmatrix} = \frac{q}{4\pi\epsilon_0 c^2}\,\frac{\omega^2 x_0 e^{i\omega(t-r/c)}}{r}\ \text{(aprox.)}. \tag{30.11}$$

Esta fórmula no es totalmente correcta porque *no* deberíamos haber usado la aceleración de la carga, sino su *componente* perpendicular a la línea QP. No obstante, supondremos que el punto P está tan lejano comparado con la distancia del punto Q al eje (la distancia ρ en la figura 30-10) de las cargas que necesitamos considerar que podemos omitir el factor coseno (que de todos modos debería ser muy cercano a 1).

Para obtener el campo total en P, sumamos ahora los efectos de todas las cargas en el plano. Deberíamos, por supuesto, hacer una suma *vectorial.* Pero ya que la dirección del campo eléctrico es casi la misma para todas las cargas, manteniendo la aproximación ya hecha, podemos sumar los módulos de los campos. En nuestra aproximación, el campo en P depende sólo de la distancia r de manera que todas las cargas a la misma distancia r producen campos iguales. Así, pues, sumamos primero los campos de todas las cargas en un anillo de ancho $d\rho$ y radio ρ. Integrando luego sobre todo ρ obtendremos el campo total.

El número de cargas en el anillo es el producto del área de la superficie del anillo $2\pi\rho d\rho$ y η, el número de cargas por unidad de área. Tenemos entonces

$$\text{Campo total en } P = \int \frac{q}{4\pi\epsilon_0 c^2}\,\frac{\omega^2 x_0 e^{i\omega(t-r/c)}}{r}\cdot\eta\cdot 2\pi\rho\,d\rho. \tag{30.12}$$

Deseamos calcular esta integral desde $p = 0$ hasta $p = \infty$. La variable t, por supuesto, debe mantenerse fija mientras efectuamos la integral de manera que las únicas cantidades variables son ρ y r. Dejando de lado por el momento todos los factores constantes, *incluyendo el factor* $e^{i\omega t}$, la integral que deseamos es

$$\int_{\rho=0}^{\rho=\infty} \frac{e^{-i\omega r/c}}{r}\,\rho\,d\rho. \tag{30.13}$$

Para realizar esta integral debemos usar la relación entre r y ρ:

$$r^2 = \rho^2 + z^2. \tag{30.14}$$

Ya que z es independiente de ρ, cuando derivamos esta ecuación obtenemos:

$$2r\,dr = 2\rho\,d\rho,$$

lo que es afortunado, porque en nuestra integral podemos remplazar $\rho d\rho$ por rdr y la r va a cancelar la del denominador. La integral que queremos es entonces de lo más simple

$$\int_{r=z}^{r=\infty} e^{-i\omega r/c}\,dr. \tag{30.15}$$

Integrar una exponencial es muy fácil. Dividimos entre el coeficiente de r en el exponente y calculamos la exponencial en los límites. Pero los límites de r no son los mismos que los límites de ρ. Cuando $\rho = 0$, tenemos $r = z$ de manera que los límites de r son de z a infinito. Obtenemos la integral:

$$-\frac{c}{i\omega}[e^{-i\infty} - e^{-(i\omega/c)z}], \tag{30.16}$$

donde escribimos ∞ en vez de $(\omega/c)\infty$ ¡ya que ambos significan un número grande!

Ahora bien, $e^{-i\infty}$ es una cantidad misteriosa. Su parte real, por ejemplo, es $\cos(-\infty)$ que matemáticamente hablando es totalmente indefinido (¡aunque esperaríamos que estuviera en alguna parte —o en todas partes (?)— entre $+1$ y -1!). Pero en una situación *física* puede significar algo muy razonable y usualmente se puede tomar como cero. Para ver que esto es así en nuestro caso, volvemos a considerar la integral original (30.15).

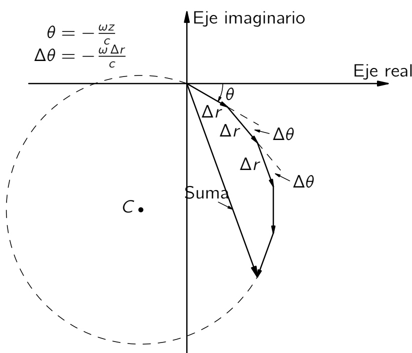


Figura 30-11. Solución gráfica de $\int_z^\infty e^{-i\omega r/c}\,dr$.

Podemos entender (30.15) como la suma de muchos números complejos pequeños, cada uno de módulo Δr y con un ángulo $\theta = -\omega r/c$ en el plano complejo. Podemos tratar de calcular la suma por un método gráfico. En la figura 30-11 hemos dibujado los primeros cinco pedazos de la suma. Cada segmento de la curva tiene un largo Δr y está colocado en un ángulo $\Delta\theta = -\omega\Delta r/c$ respecto al pedazo anterior. La suma de estos cinco primeros pedazos está representada por la flecha desde el punto de partida hasta el extremo del quinto segmento. A medida que seguimos sumando pedazos, vamos a trazar un polígono hasta que volvemos al punto inicial (aproximadamente) y después vamos a empezar la vuelta de nuevo. Al sumar más pedazos sólo damos vuelta y vuelta, manteniéndonos cerca de un círculo cuyo radio se demuestra fácilmente ser c/ω. ¡Ahora podemos ver por qué la integral no da una respuesta definida!

Pero ahora debemos volver a la *física* de la situación. En una situación real el plano de cargas *no puede* ser de extensión infinita, sino que debe terminar alguna vez. Si terminara súbitamente y fuera exactamente de forma circular, nuestra integral tendría algún valor en el círculo de la figura 30-11. Sin embargo, si dejamos que el número de cargas en el plano vaya disminuyendo gradualmente a una cierta distancia grande del centro (o bien que termine de repente, pero en una forma irregular de manera que para ρ grande todo el anillo de ancho $d\rho$ no contribuya más), entonces el coeficiente η en la integral exacta va a decrecer hasta cero. Ya que estamos sumando pedazos más pequeños, pero aún girando en el mismo ángulo, el gráfico de nuestra integral resultará ser entonces una curva que es una espiral. La espiral terminará finalmente en el centro de nuestro círculo original como se muestra en la figura 30-12. La integral *físicamente* correcta es el número complejo A representado en la figura por el intervalo desde el punto inicial al centro del círculo que es igual a

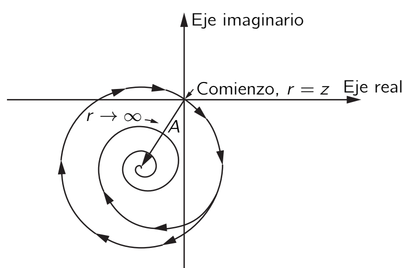


Figura 30-12. Solución gráfica de $\int_z^\infty \eta e^{-i\omega r/c}\,dr$.

$$\frac{c}{i\omega}\,e^{-i\omega z/c}, \tag{30.17}$$

como ustedes mismos pueden desarrollar. Éste es el mismo resultado que obtendríamos de la ecuación (30.16) si hacemos $e^{-i\infty} = 0$.

(Existe además una razón por qué la contribución a la integral disminuye para valores grandes de r y ése es el factor que hemos omitido para la proyección de la aceleración en el plano perpendicular a la línea PQ.)

Estamos interesados, por supuesto, sólo en situaciones físicas, así que vamos a tomar $e^{-i\infty}$ igual a cero. Volviendo a nuestra fórmula original (30.12) para el campo y volviendo a poner todos los factores que acompañan la integral, tenemos el resultado

$$\text{Campo total en } P = -\frac{\eta q}{2\epsilon_0 c}\,i\omega x_0 e^{i\omega(t-z/c)} \tag{30.18}$$

(recordando que $1/i = -i$).

Es interesante notar que $(i\omega x_0 e^{i\omega t})$ es igual a la *velocidad* de las cargas, así que también podemos escribir la ecuación del campo como:

$$\text{Campo total en } P = -\frac{\eta q}{2\epsilon_0 c}\,(\text{velocidad de las cargas})_{\text{en } t\,-\,z/c}, \tag{30.19}$$

lo que es un poco extraño, porque el retardo es justamente en la distancia z, que es la distancia más corta desde P al plano de cargas. Pero así es como resulta, afortunadamente, una fórmula bastante simple. (Podemos agregar, entre paréntesis, que aunque nuestra derivación es válida sólo para distancias alejadas del plano de cargas oscilantes, resulta que la fórmula (30.18) o (30.19) es correcta para cualquier distancia, aún para $z < \lambda$.)

31

El origen del índice de refracción

31-1 El índice de refracción

Hemos dicho antes que la luz va más lenta en el agua que en el aire y ligeramente más lenta en el aire que en el vacío. Este efecto está descrito por el índice de refracción n. Ahora nos gustaría entender cómo se origina esta velocidad más baja. En particular, deberíamos tratar de ver qué relación hay con algunas hipótesis físicas o afirmaciones que hicimos anteriormente y que eran las siguientes:

(a) Que el campo eléctrico total, en cualquier circunstancia física, siempre se puede representar por la suma de los campos de todas las cargas del universo.

(b) Que el campo de una sola carga está dado *siempre* por su aceleración calculada con un retardo a la velocidad c (para el campo de *radiación*).

Pero para un trozo de vidrio ustedes podrían pensar: "Ah, no, usted debería modificar todo esto. Debería decir que está retardada a la velocidad c/n". Eso, sin embargo, no es correcto y debemos entender por qué no lo es.

Es aproximadamente verdadero que la luz o cualquier onda eléctrica *parezca* viajar a la velocidad c/n a través de un material cuyo índice de refracción es n. Pero, sin embargo, los campos son producidos por los movimientos de *todas* las cargas —incluyendo las cargas que se mueven en el material— y con estas contribuciones básicas del campo que en última instancia viaja a la velocidad c, nuestro problema es entender cómo se origina la velocidad *aparentemente* más baja.

Intentaremos entender el efecto en un caso muy simple. Una fuente que nosotros debemos llamar "la fuente *externa*" se coloca a una gran distancia de una placa delgada de material transparente, digamos vidrio. Preguntamos por el campo que está a gran distancia en el lado opuesto de la placa. La situación se encuentra ilustrada por el diagrama de la figura 31-1, donde se pensará que S y P están muy lejos de la placa. De acuerdo con los principios que hemos establecido anteriormente, un campo eléctrico en cualquier parte que esté, lejos de todas las cargas en movimiento, es la suma (vectorial) de los campos producidos por la fuente externa (en S) *y* los campos producidos por *cada una* de las cargas a la placa de vidrio, *cada uno con su retardo propio con velocidad c*. Recuerden que la contribución de cada carga no cambia por la presencia de otras cargas. Éstos son nuestros principios básicos. El campo en P se puede escribir así:

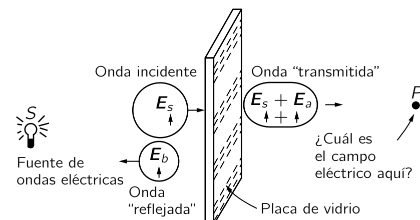


Figura 31-1. Ondas eléctricas atraviesan una capa de material transparente.

$$\boldsymbol{E} = \sum_{\text{todas las cargas}} \boldsymbol{E}_{\text{cada carga}} \tag{31.1}$$

o bien,

$$\boldsymbol{E} = \boldsymbol{E}_s + \sum_{\text{todas las otras cargas}} \boldsymbol{E}_{\text{cada carga}}, \tag{31.2}$$

donde $\boldsymbol{E}_s$ es el campo debido a la fuente sola y sería precisamente el campo P *si no hubiera material presente*. Esperamos que el campo en P sea diferente de $\boldsymbol{E}_s$ si hay otras cargas en movimiento.

¿Por qué habría de haber cargas moviéndose en el vidrio? Sabemos que todo material consiste en átomos que contienen electrones. Cuando el campo eléctrico

de la fuente actúa sobre estos átomos mueve los electrones de arriba abajo porque ejerce una fuerza sobre ellos. Y los electrones en movimiento generan un campo —constituyen nuevos radiadores—. Estos nuevos radiadores están relacionados con la fuente S, porque están accionados por el campo de la fuente. El campo total no es exactamente el campo de la fuente S, sino que está modificado por la contribución adicional de otras cargas en movimiento. Esto significa que el campo no es el mismo que el que había antes que el vidrio se colocara ahí, sino que se modifica y resulta que se modifica de tal manera que el campo dentro del vidrio pareciera estar moviéndose a una velocidad diferente. Ésta es la idea que nos gustaría desarrollar cuantitativamente.

Ahora bien, esto es bien complicado en el caso exacto, porque a pesar de haber dicho que todas las otras cargas en movimiento son accionadas por el campo de la fuente, eso no es totalmente cierto. Si consideramos una carga en particular, no sólo siente la fuente, sino que, como cualquier otra cosa en el mundo, siente *todas* las cargas que se están moviendo. Siente, en particular, las cargas que se están moviendo en cualquier otro lugar del vidrio. Por consiguiente, el campo total que está actuando en una *carga particular* es una combinación de los campos de las otras cargas *¡cuyos movimientos dependen de lo que esté haciendo esta carga particular!* Pueden apreciar que se necesitaría un complicado conjunto de ecuaciones para obtener la fórmula completa y exacta. Es tan complicado que postergamos este problema hasta el próximo año.

En cambio, desarrollaremos un caso muy simple para entender muy claramente los principios físicos. Tomamos un caso en el cual los efectos de los otros átomos son muy pequeños comparados con los efectos de la fuente. En otras palabras, tomamos un material en el cual el campo total no se modifica mucho por el movimiento de las otras cargas. Eso corresponde a un material cuyo índice de refracción está muy cerca de 1, lo que sucederá, por ejemplo, si la densidad de los átomos es muy baja. Nuestro cálculo va a ser válido para cualquier caso en el cual el índice está, por cualquier razón, muy cerca de 1. De esta manera evitaremos las complicaciones de la solución completa más general.

A propósito, notemos que hay otro efecto causado por el movimiento de las cargas en la placa. Estas cargas también van a irradiar ondas para atrás hacia la fuente S. Este campo que va hacia atrás es la luz que vemos reflejada por las superficies de materiales transparentes. No proviene solamente de la superficie. La radiación hacia atrás proviene de todas partes del interior, pero resulta que el efecto total es equivalente a una reflexión desde las superficies. Así, estos efectos de reflexión están más allá de nuestra aproximación porque estaremos limitados a un cálculo para un material cuyo índice está tan cerca de 1 que se refleja muy poca luz.

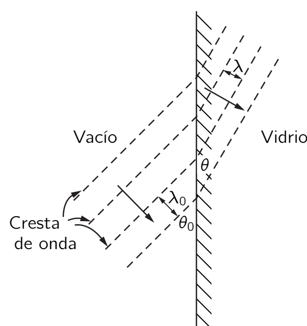


Figura 31-2. Relación entre refracción y cambio de velocidad.

Antes que prosigamos con nuestro estudio de cómo se genera el índice de refracción, debemos saber que todo lo que se necesita para entender la refracción es comprender por qué la *velocidad* de onda aparente es diferente en materiales distintos. La *desviación* de los rayos de luz se produce *porque* la rapidez efectiva de las ondas es diferente en cada material. Para recordarles cómo sucede eso, hemos dibujado en la figura 31-2 varias crestas sucesivas de una onda eléctrica que llega desde el vacío a la superficie de un bloque de vidrio. La flecha perpendicular a las crestas de la onda indica la dirección de movimiento de la onda. Todas las oscilaciones de la onda deben tener ahora igual *frecuencia*. (Vimos que las oscilaciones forzadas tienen la misma frecuencia que la fuente excitadora.) Esto significa, también, que las crestas de las ondas para las ondas en ambos lados de la superficie deben tener la *misma separación a lo largo de la superficie* porque deben viajar juntas, de manera que una carga situada en la separación perciba una sola frecuencia. La distancia *más corta* entre crestas de onda a onda es la longitud de onda, que es la velocidad dividida entre la frecuencia. En el lado del vacío es $\lambda_0 = 2\pi c/\omega$ y en el otro lado es $\lambda = 2\pi v/\omega$, o bien, $2\pi c/\omega n$ si $v = c/n$ es la velocidad de onda. A partir de la figura podemos ver que la única manera de que las ondas "encajen" apropiadamente en la superficie de separación es que las

ondas en el material estén viajando a un ángulo diferente respecto a la superficie. Por la geometría de la figura pueden ver que para un "encaje" debemos tener $\lambda_0/\operatorname{sen}\theta_0 = \lambda/\operatorname{sen}\theta$, o bien, $\operatorname{sen}\theta_0/\operatorname{sen}\theta = n$, que es la leyes de Snell. Vamos a considerar, para el resto de nuestra discusión, sólo por qué la luz tiene una velocidad efectiva c/n en un material de índice n y no nos preocuparemos más en este capítulo de la desviación de la luz.

Volvamos ahora a la situación que se muestra en la figura 31-1. Vemos que lo que tenemos que hacer es calcular el campo producido en P por todas las cargas oscilantes en la placa de vidrio. Vamos a llamar E_a esta parte del campo eléctrico y ésta es justamente la suma escrita como segundo término en la ecuación (31.2). Cuando sumamos al término E_s debido a la fuente, vamos a tener el campo total en P.

Esto es probablemente lo más difícil que haremos este año, pero es complicado sólo en que hay muchas partes que deben juntarse; cada parte, sin embargo, es muy simple. A diferencia de otras deducciones, donde decimos: "olviden la deducción, miren sólo el resultado", en este caso no necesitamos tanto el resultado como la deducción. En otras palabras, lo que debemos entender ahora es el mecanismo físico para la producción del índice.

Para ver a dónde vamos, averiguamos primero cuánto debería ser el "campo correctivo" E_a, si el campo total en P se va a parecer a una radiación desde la fuente que se frena a medida que pasa por una placa delgada. Si la lámina no tuviera efecto en él, el campo de una onda que se propaga hacia la derecha (según el eje z) sería:

$$E_s = E_0 \cos\omega(t - z/c) \tag{31.3}$$

o bien, usando la notación exponencial,

$$E_s = E_0 e^{i\omega(t-z/c)}. \tag{31.4}$$

Ahora bien, ¿qué pasaría si la onda se propagara más lentamente al ir a través de la placa? Llamemos Δz, el espesor de la placa. Si la placa no estuviera ahí, la onda atravesaría la distancia Δz en un tiempo $\Delta z/c$. Pero si se propagara con velocidad c/n, entonces deberá demorarse un tiempo mayor $n\Delta z/c$ o el tiempo *adicional* $\Delta t = (n-1)\Delta z/c$. Después de eso seguiría propagándose nuevamente con velocidad c. Podemos tomar en cuenta el atraso adicional al pasar por la placa reemplazando t en la ecuación (31.4) por $(t - \Delta t)$ o por $\big[t - (n-1)\Delta z/c\big]$. Por consiguiente, después de intercalar la placa la onda debería escribirse:

$$E_{\text{después de la placa}} = E_0 e^{i\omega[t-(n-1)\,\Delta z/c - z/c]}. \tag{31.5}$$

También podemos escribir esta ecuación como:

$$E_{\text{después de la placa}} = e^{-i\omega(n-1)\,\Delta z/c} E_0 e^{i\omega(t-z/c)}, \tag{31.6}$$

que dice que la onda después de la placa se obtiene de la onda que podría existir sin la placa, es decir, E_s multiplicado por el factor $e^{-i\omega(n-1)\Delta z/c}$. Ahora sabemos que multiplicar una función oscilante como $e^{i\omega t}$ por un factor $e^{i\theta}$ sólo significa cambiar la fase de oscilación en un ángulo de θ, que es, por supuesto, lo que ha hecho el atraso extra al pasar por un espesor Δz. Ha retardado la fase en una cantidad $\omega(n-1)\Delta z/c$ (retardado debido al signo menos en el exponente).

Hemos dicho anteriormente que la placa debiera *sumar* un campo E_a al campo original $E_s = E_0 e^{i\omega(t-z/c)}$, pero hemos encontrado, en cambio, que el efecto de la placa es *multiplicar* el campo por un factor que cambia su fase. Sin embargo, eso está en realidad muy bien, porque podemos obtener el mismo resultado sumando un número complejo apropiado. En particular es fácil encontrar el número correcto al sumar, en el caso que Δz sea pequeño, porque deben recordar que si x es un número pequeño entonces e^x es casi igual a $(1 + x)$. Podemos escribir, por lo tanto,

$$e^{-i\omega(n-1)\,\Delta z/c} \approx 1 - i\omega(n-1)\,\Delta z/c. \tag{31.7}$$

Usando esta aproximación en la ecuación (31.6) obtenemos:

$$E_{\text{después de la placa}} = \underbrace{E_0 e^{i\omega(t-z/c)}}_{E_s} - \underbrace{\frac{i\omega(n-1)\,\Delta z}{c}\,E_0 e^{i\omega(t-z/c)}}_{E_a}. \tag{31.8}$$

El primer término es justamente el campo de la fuente y el segundo término debe ser, justamente, igual a E_a, el campo producido a la derecha de la placa por las cargas oscilantes de la placa, expresado aquí en términos del índice de refracción n y dependiendo, por supuesto, de la intensidad de la onda de la fuente.

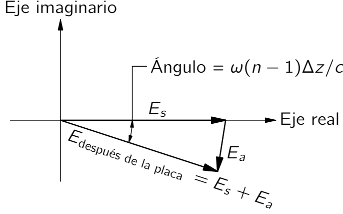


Figura 31-3. Diagrama para la onda transmitida para un t y z particular.

Lo que hemos estado haciendo se visualiza fácilmente si observamos el diagrama de números complejos en la figura 31-3. Primero dibujamos el número E_s (elegimos algunos valores de z y t de manera que E_s resulte horizontal, pero esto no es necesario). El atraso debido a la reducción de la velocidad en la placa retrasará la fase de este número, o sea, hará rotar E_s en un ángulo negativo. Pero esto es equivalente a sumar el pequeño vector E_a a un ángulo casi recto con E_s. Pero esto es justamente lo que significa el factor $-i$ en el segundo término de la ecuación (31.8). Ésta dice que si E_s es real, entonces E_a es imaginario negativo o que, en general, E_s y E_a forman un ángulo recto.

31-2 El campo debido al medio

Ahora debemos preguntar: ¿es el campo E_a obtenido en el segundo término de la ecuación (31.8) del tipo que uno esperaría de cargas oscilantes en la placa? Si podemos demostrar que lo es, ¡entonces habremos calculado lo que debe ser el índice n!, ya que n es el único número que no es fundamental en la ecuación (31.8). Calculemos ahora qué campo E_a van a producir las cargas en el medio (para ayudarles a estar al tanto de los muchos símbolos que hemos usado hasta ahora y que estaremos usando durante el resto de nuestro cálculo, los hemos reunido en la tabla 31-1).

Tabla 31-1

Símbolos usados en los cálculos

E_s =	campo de la fuente
E_a =	campo producido por las cargas en la placa
Δz =	espesor de la placa
z =	distancia perpendicular desde la placa
n =	índice de refracción
ω =	frecuencia (angular) de la radiación
N =	número de cargas por unidad de volumen en la placa
η =	número de cargas por unidad de superficie de la placa
q_e =	carga de un electrón
m =	masa de un electrón
ω_0 =	frecuencia de resonancia de un electrón ligado a un átomo

Si la fuente S (de la figura 31-1) está lejos a la izquierda, entonces el campo E_s tendrá la misma fase en toda la placa, de manera que podemos escribir que en la vecindad de la placa es

$$E_s = E_0 e^{i\omega(t-z/c)}. \tag{31.9}$$

Justo en la placa, donde $z = 0$, vamos a tener:

$$E_s = E_0 e^{i\omega t} \text{ (en la placa).} \tag{31.10}$$

Cada uno de los electrones de los átomos de la placa percibirá este campo eléctrico y será forzado hacia arriba y abajo (suponemos que la dirección de E_0 es vertical) por la fuerza eléctrica qE. Para encontrar qué movimiento esperamos de los electrones supondremos que los átomos son pequeños osciladores, esto es, que los electrones están sujetos elásticamente a los átomos, lo que significa que si se aplica una fuerza a un electrón, su desplazamiento de la posición normal será proporcional a la fuerza.

Ustedes pueden pensar que éste es un modelo raro de un átomo si han oído de electrones que giran en órbitas. Pero ése es sólo una imagen muy simplificada. La imagen correcta de un átomo que está dada por la teoría de la mecánica ondulatoria indica que, *en lo que se refiere a los problemas que incluyen a la luz*, los electrones se comportan como si estuvieran sujetos por resortes. Así, supondremos que los electrones tienen una fuerza restauradora lineal que junto con su masa m los hace comportarse como pequeños osciladores con una frecuencia de resonancia ω_0. Ya hemos estudiado tales osciladores y sabemos que la ecuación de su movimiento se escribe así:

$$m\left(\frac{d^2x}{dt^2} + \omega_0^2 x\right) = F, \tag{31.11}$$

donde F es la fuerza impulsora.

Para nuestro problema la fuerza impulsora proviene del campo eléctrico de la onda de la fuente, por lo tanto usamos

$$F = q_e E_s = q_e E_0 e^{i\omega t}, \tag{31.12}$$

donde q_e es la carga eléctrica del electrón y para E_s usamos la expresión $E_s = E_0 e^{i\omega t}$ de (31.10). Nuestra ecuación de movimiento para el electrón es entonces

$$m\left(\frac{d^2x}{dt^2} + \omega_0^2 x\right) = q_e E_0 e^{i\omega t}. \tag{31.13}$$

Ya hemos resuelto esta ecuación anteriormente y sabemos que la solución es:

$$x = x_0 e^{i\omega t}, \tag{31.14}$$

donde, sustituyendo en (31.13), encontramos que

$$x_0 = \frac{q_e E_0}{m(\omega_0^2 - \omega^2)}, \tag{31.15}$$

de manera que

$$x = \frac{q_e E_0}{m(\omega_0^2 - \omega^2)}\, e^{i\omega t}. \tag{31.16}$$

Tenemos lo que necesitábamos conocer —el movimiento de los electrones en la placa—. Y es el mismo para cada electrón, excepto que la posición media (el "cero" del movimiento) es, por supuesto, diferente para cada electrón.

Ahora estamos listos para encontrar el campo E_a que estos átomos producen en el punto P, porque ya hemos encontrado (al final del capítulo 30) el campo producido por una hoja de cargas que se mueven en conjunto. Refiriéndonos de nuevo a la ecuación (30.19) vemos que el campo E_a en P es justamente una constante negativa por la velocidad de las cargas retardadas en el tiempo en la cantidad z/c. Derivando x en la ecuación (31.16) para obtener la velocidad e introduciendo el retardo, o simplemente introduciendo x_0 de (31.15) en (30.18), resulta:

$$E_a = -\frac{\eta q_e}{2\epsilon_0 c}\left[i\omega\, \frac{q_e E_0}{m(\omega_0^2 - \omega^2)}\, e^{i\omega(t-z/c)}\right]. \tag{31.17}$$

Tal como esperábamos, el movimiento forzado de los electrones produjo una onda adicional que viaja hacia la derecha (eso es lo que dice el factor $e^{i\omega(t-z/c)}$) y

la amplitud de esta onda es proporcional al número de átomos por unidad de superficie en la placa (el factor η) y también proporcional a la intensidad del campo de la fuente (el factor E_0). Luego hay algunos factores que dependen de las propiedades atómicas (q_e, m y ω_0) como habíamos esperado.

Lo más importante, sin embargo, es que esta fórmula (31.17) para E_a se parece mucho a la expresión para E_a que obtuvimos en la ecuación (31.8), diciendo que la onda original se retrasó al pasar por un material con un índice de refracción n. Las dos expresiones serán, en efecto, idénticas si

$$(n-1)\,\Delta z = \frac{\eta q_e^2}{2\epsilon_0 m(\omega_0^2 - \omega^2)}. \tag{31.18}$$

Noten que ambos lados son proporcionales a Δz, ya que η, que es el número de átomos *por unidad de área*, es igual a $N\Delta z$, donde N es el número de átomos *por unidad de volumen* de la placa. Sustituyendo $N\Delta z$ por η y simplificando los Δz obtenemos nuestro resultado principal, una fórmula para el índice de refracción en términos de las propiedades de los átomos del material y de la frecuencia de la luz:

$$n = 1 + \frac{N q_e^2}{2\epsilon_0 m(\omega_0^2 - \omega^2)}. \tag{31.19}$$

Esta ecuación da la "explicación" del índice de refracción que queríamos obtener.

31-3 Dispersión

Noten que en el proceso anterior hemos obtenido algo muy interesante. Porque no sólo tenemos un número para el índice de refracción que puede calcularse a partir de las cantidades atómicas básicas, sino que también hemos aprendido cómo varía el índice de refracción con la frecuencia ω de la luz. Esto es algo que no entenderíamos nunca de la simple afirmación de que "la luz se propaga más lentamente en un material transparente". Todavía tenemos el problema, por cierto, de saber cuántos átomos hay por unidad de volumen y cuál es su frecuencia ω_0. No sabemos esto todavía, porque es diferente para cada material y no podemos obtener una teoría general de eso ahora. La formulación de una teoría general de las propiedades de diferentes sustancias —sus frecuencias naturales, etc.— es posible sólo con la mecánica atómica cuántica. Además, materiales diferentes tienen distintas propiedades e índices diferentes; por lo tanto, no podemos esperar obtener una fórmula general para el índice que se aplique a todas las sustancias.

Sin embargo, discutiremos la fórmula que hemos obtenido, en varios casos posibles. Ante todo, para la mayoría de los gases ordinarios (por ejemplo, aire, la mayoría de los gases incoloros, hidrógeno, helio, etc.) las frecuencias naturales de los osciladores electrónicos corresponden a la luz ultravioleta. Estas frecuencias son mayores que las frecuencias de la luz visible, o sea, ω_0 es mucho mayor que el ω de la luz visible y en una primera aproximación podemos despreciar ω^2 en comparación con ω_0^2. Entonces encontramos que el índice es casi constante. Por consiguiente, para un gas el índice es casi constante. Esto también es cierto para la mayoría de otras sustancias transparentes, como el vidrio. Sin embargo, si observamos nuestra expresión más de cerca, notamos que a medida que ω aumenta, disminuyendo así un poquito el denominador, el índice también aumenta. Por consiguiente, n aumenta lentamente con la frecuencia. El índice es más grande para la luz azul que para la luz roja. Ésa es la razón por la cual un prisma desvía más la luz en el azul que en el rojo.

El fenómeno que el índice depende de la frecuencia se llama fenómeno de *dispersión** porque es la base del hecho que la luz sea "dispersada" en un espectro por un prisma. La ecuación para el índice de refracción en función de la frecuencia se llama *ecuación de dispersión*. Hemos obtenido así una ecuación de Normal.

* N. del T.: No confundir esta propiedad de un medio con el proceso de disprsión anómala de la luz por centros dispersos (véase la sección 32-5).

(En los últimos años, las “ecuaciones de dispersión” han encontrado un nuevo uso en la teoría de las partículas elementales.)

Nuestra ecuación de dispersión sugiere otros efectos interesantes. Si tenemos una frecuencia natural ω_0, que está en la región visible, o si medimos el índice de refracción de un material como vidrio en el ultravioleta, donde ω se acerca a ω_0, vemos que a frecuencias muy cercanas a la frecuencia natural el índice se puede hacer enormemente grande, porque el denominador puede hacerse cero. A continuación, supongan que ω es mayor que ω_0. Esto sucedería, por ejemplo, si tomáramos un material como vidrio y lo iluimináramos con radiación de rayos X. En efecto, ya que muchos materiales que son opacos a la luz visible, como el grafito por ejemplo, son transparentes a los rayos, también podemos hablar del índice de refracción del carbono para los rayos X. Todas las frecuencias naturales de los átomos de carbón son mucho más bajas que la frecuencia que estamos usando en los rayos X ya que la radiación de los rayos X tiene una frecuencia muy alta. El índice de refracción es el dado por nuestra ecuación de dispersión, si hacemos ω_0, igual a cero (despreciamos w_0^2 comparado con w^2).

Una situación similar ocurriría si dirigiéramos ondas de radio (o luz) sobre un gas de electrones libres. En la atmósfera superior, los electrones son liberados de sus átomos por luz ultravioleta del Sol y permanecen ahí como electrones libres. Para electrones libres $\omega_0 = 0$ (no hay fuerza elástica restauradora). Haciendo $\omega_0 = 0$ en nuestra ecuación de dispersión se obtiene la fórmula correcta para el índice de refracción de ondas de radio en la estratosfera, donde N representa ahora la densidad de electrones libres (número por unidad de volumen) en la estratosfera. Pero observemos nuevamente la ecuación, si hacemos incidir rayos X sobre la materia u ondas de radio (o cualquier onda eléctrica) sobre electrones libres, el término $(w_0^2 - w^2)$ se hace *negativo* y obtenemos el resultado que n es *menor que uno.* ¡Esto significa que la velocidad efectiva de las ondas en la sustancia es *mayor* que c! ¿Puede esto ser correcto?

Es correcto. A pesar de que se dice que no se pueden mandar señales a velocidad mayor que la velocidad de la luz, es, sin embargo, cierto que el índice de refracción de materiales de una cierta frecuencia particular puede ser mayor o menor que 1. Esto significa sólo que el *cambio de fase* que es producido por la luz dispersada puede ser positivo o negativo. Se puede demostrar, sin embargo, que la velocidad a la cual se puede mandar una *señal* no está determinada por el índice a una frecuencia, sino que depende de lo que es el índice a *diferentes* frecuencias. Lo que el índice nos indica es la velocidad a la cual se propagan los *nodos* (o crestas) de las ondas. El *nodo* de una onda no es una señal por sí mismo. Es una onda perfecta, que no tiene modulaciones de ningún tipo, es decir, que es una oscilación estacionaria, no se puede decir realmente cuándo “empieza”, así que no se puede usar como señal para medir tiempo. Para enviar una *señal* se tiene que cambiar la onda de alguna manera, hacer una ranura en ella, hacerla un poco más ancha o delgada. Eso significa que se debe tener más de una frecuencia en la onda y se puede demostrar que la velocidad a la cual se propagan las *señales* no depende sólo del índice, sino de la manera como el índice cambia con la frecuencia. Este tema también tenemos que postergarlo (hasta el capítulo 48). Entonces calcularemos para ustedes la velocidad real de las *señales* a través de un pedazo de vidrio y verán que no será mayor que la velocidad de la luz, aunque los nodos, que son puntos matemáticos, se propagan más rápidamente que la luz.

Sólo para dar una pequeña indicación de cómo sucede eso, ustedes notarán que la verdadera dificultad tiene que ver con el hecho de que las respuestas de las cargas son opuestas al campo, es decir, la señal se ha invertido. Así, en nuestra expresión para x, en la ecuación (31.16), el desplazamiento de la carga tiene dirección opuesta a la del campo impulsor, porque $(w_0^2 - w^2)$ es negativo para w_0 pequeño. La fórmula dice que cuando el campo eléctrico está tirando en una dirección, la carga se está moviendo en la dirección opuesta.

¿Cómo sucede que la carga esté yendo en la dirección opuesta? Ciertamente que no parte en la dirección opuesta cuando se aplica el campo al comienzo. Cuando el movimiento recién empieza hay un transitorio que se extingue después

de un rato y sólo *entonces* está la fase de oscilación de la carga opuesta a la del campo impulsor. Y es entonces que la *fase* del campo transmitido puede aparecer como *adelantada* respecto a la onda de la fuente. Es a este *adelanto de fase* que nos referimos cuando decimos que la “velocidad de fase” o velocidad de los nodos es mayor que c. En la figura 31-4 damos una idea esquemática de cómo se verían las ondas para el caso en que la onda se pusiera en marcha súbitamente (para hacer una señal). Van a ver en el diagrama que la *señal* (es decir, el *comienzo* de la onda) *no* está primero para la onda que termina con un adelanto de fase.

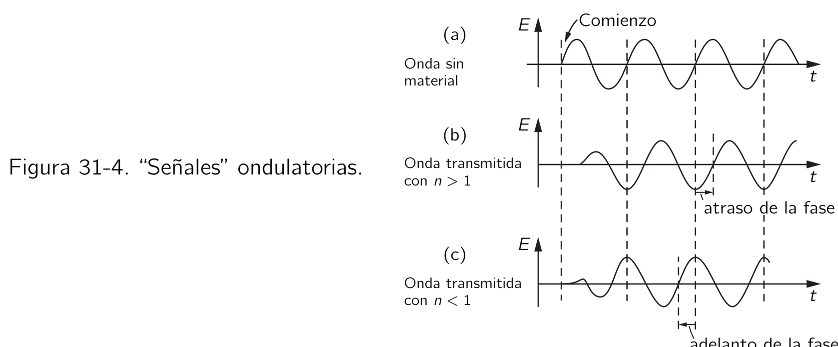


Figura 31-4. “Señales” ondulatorias.

Observemos de nuevo nuestra ecuación de dispersión. Debemos hacer notar que nuestro análisis de refracción da un resultado que es algo más simple de lo que realmente se encuentra en la naturaleza. Para ser totalmente precisos, debemos agregar algunos refinamientos. Primero, deberíamos esperar que nuestro modelo de oscilador atómico tuviera alguna fuerza de amortiguamiento (de otra manera, una vez puesto en movimiento, oscilaría para siempre y no podemos esperar que esto suceda). Hemos desarrollado antes, ecuación (23.8), el movimiento de un oscilador amortiguado y el resultado es que el denominador en la ecuación (31.16) y, por lo tanto, en (31.19) se cambia de $(\omega_0^2 - \omega^2)$ a $(\omega_0^2 - \omega^2 + i\gamma\omega)$, donde γ es el coeficiente de amortiguamiento.

Necesitamos una segunda modificación para tener en cuenta el hecho que hay varias frecuencias de resonancia para un tipo particular de átomo. Es fácil arreglar nuestra ecuación de dispersión, imaginando que hay varios tipos diferentes de osciladores, pero que cada oscilador actúa separadamente y por lo tanto simplemente sumamos las contribuciones de todos los osciladores. Digamos que hay N_k electrones por unidad de volumen cuya frecuencia natural es ω_k y cuyo factor de amortiguamiento es γ_k. Entonces tendríamos para nuestra ecuación de dispersión:

$$n = 1 + \frac{q_e^2}{2\epsilon_0 m} \sum_k \frac{N_k}{\omega_k^2 - \omega^2 + i\gamma_k\omega}. \tag{31.20}$$

Tenemos finalmente una expresión completa que describe el índice de refracción observado para muchas sustancias.* El índice descrito por esta fórmula varía con la frecuencia aproximadamente como en la curva mostrada en la figura 31-5.

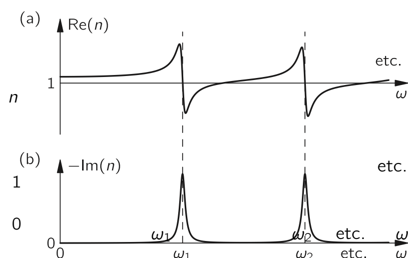


Figura 31-5. El índice de refracción en función de la frecuencia.

Van a notar que mientras ω no esté demasiado cerca a una de las frecuencias de resonancia, la pendiente de la curva es positiva. Tal pendiente positiva se llama dispersión ‘normal” (porque claramente es la que se presenta más comúnmente). Muy cerca de las frecuencias de resonancia, sin embargo, hay un pequeño intervalo de los ω para los cuales la pendiente es negativa. Tal pendiente negativa se le llama a menudo dispersión “anómala”, porque parecía poco usual cuando se observó por primera vez mucho antes de que nadie ni siquiera supiera que existían

* En realidad, aunque en mecánica cuántica la ecuación (31.20) aún es válida, su interpretación es algo diferente. En mecánica cuántica aun un átomo con un electrón, como el hidrógeno, tiene varias frecuencias de resonancia. Por lo tanto, N_k no es realmente el número de electrones con frecuencia ω_k, sino que se remplaza por Nf_k, donde N es el número de átomos por unidad de volumen y f_k (llamado la intensidad del oscilador) es un factor que indica lo fuertemente que exhiben los átomos cada una de sus frecuencias de resonancia ω_k.

cosas como los electrones. ¡Desde nuestro punto de vista, ambas pendientes son perfectamente "normales"!

31-4 Absorción

Quizás ustedes se hayan percatado de algo un poco extraño acerca de la última forma, ecuación (31.20), que obtuvimos para nuestra ecuación de dispersión. Debido al término $i\gamma$ que introdujimos para tener en cuenta la amortiguación ¡el índice de refracción es ahora un *número complejo*! ¿Qué significa *eso*? Calculando las partes real e imaginaria de n podríamos escribir

$$n = n' - in'', \tag{31.21}$$

donde n' y n'' son números reales. (Usamos el signo menos frente a in'', porque entonces n'' resultará un número positivo, como ustedes mismos pueden demostrar.)

Podemos ver lo que significa este índice complejo cuando hay una sola frecuencia de resonancia volviendo a la ecuación (31.6), que es la ecuación de la onda después de que pasa por una placa de material con índice n. Si ponemos nuestro número complejo n en esta ecuación y reordenamos, obtenemos

$$E_{\text{después de la placa}} = \underbrace{e^{-\omega n''\,\Delta z/c}}_{\text{A}}\,\underbrace{e^{-i\omega(n'-1)\,\Delta z/c}E_0 e^{i\omega(t-z/c)}}_{\text{B}}. \tag{31.22}$$

Los últimos factores, llamados B en la ecuación (31.22), son de la forma que teníamos antes y de nuevo describen una onda cuya fase ha sido atrasada en el ángulo $\omega(n'-1)\Delta z/c$ al atravesar el material. El primer término (A) es nuevo y es un factor exponencial con exponente *real*, porque había dos i que se simplificaron. Además, el exponente es negativo, de manera que el factor es un número real menor que uno. Describe una *disminución* del módulo del campo y, como deberíamos esperar, en una cantidad que es mayor mientras mayor es Δz. A medida que la onda pasa a través del material se debilita. El material está "absorbiendo" parte de la onda. La onda sale al otro lado con menos energía. Esto no debe sorprendernos porque la amortiguación que introdujimos para los osciladores es en realidad una fuerza de fricción y debe esperarse que cause pérdida de energía. Vemos que la parte imaginaria n'' de un índice de refracción complejo representa una absorción (o "atenuación") de la onda. De hecho, n'' a veces se llama "índice de absorción".

También podríamos indicar que una parte imaginaria del índice n corresponde a inclinar la flecha E_a en la figura 31-3 hacia el origen. Está claro por qué el campo transmitido está entonces disminuido.

Normalmente, como en el vidrio, por ejemplo, la absorción de luz es muy pequeña. Esto era de esperar a partir de nuestra ecuación, (31.20), porque la parte imaginaria del denominador, $i\gamma_k\omega$, es mucho menor que el término $(\omega_k^2-\omega^2)$. Pero si la frecuencia ω de la luz está muy cercana a ω_k, entonces el término de resonancia $(\omega_k^2-\omega^2)$ puede hacerse pequeño comparado con $i\gamma_k\omega$ y el índice se hace casi completamente imaginario. La absorción de la luz llega a ser el efecto dominante. Es justamente este efecto que da las líneas negras en el espectro de la luz que recibimos del Sol. La luz de la superficie solar ha pasado a través de la atmósfera del Sol (como también la de la Tierra) y la luz ha sido absorbida fuertemente a la frecuencia de resonancia de los átomos en la atmósfera solar.

La observación de tales líneas espectrales en la luz solar nos permite conocer las frecuencias de resonancias de los átomos y, por lo tanto, la composición química de la atmósfera del Sol. El mismo tipo de observacion se tiene en los materiales de las estrellas. Por tales medidas sabemos que los elementos químicos en el Sol y en las estrellas son los mismos que encontramos en la Tierra.

31-5 La energía transportada por una onda eléctrica

Hemos visto que la parte imaginaria del índice significa absorción. Usaremos ahora este conocimiento para encontrar cuánta energía transporta una onda luminosa. Hemos dicho anteriormente que la energía transportada por la luz es proporcional a $\overline{E^2}$, el promedio temporal del cuadrado del campo eléctrico de la onda. La disminución de E debida a la absorción debe significar una pérdida de energía que se podría atribuir a cierto roce entre los electrones y podríamos adivinar que terminaría como calor en el material.

Si consideramos la luz que llega por unidad de área, digamos un centímetro cuadrado de nuestra placa en la figura 31-1, entonces podemos escribir la siguiente ecuación de energía (¡si suponemos, como lo estamos *haciendo*, que la energía se conserva!):

$$\text{Energía entrante / s} = \text{energía saliente / s} + \text{trabajo realizado / s} \tag{31.23}$$

Para el primer término podemos escribir $\alpha\overline{E_s^2}$, donde α es la constante de proporcionalidad, aún desconocida, que relaciona el valor promedio de E^2 con la energía transportada. Para el segundo término debemos incluir la parte de los átomos del material que están irradiando, así que debemos usar $\alpha\overline{(E_s + E_a)^2}$ o (calculando el cuadrado) $\alpha(\overline{E_s^2} + 2\overline{E_sE_a} + \overline{E_a^2})$.

Todos nuestros cálculos han sido hechos para una capa delgada de material, cuyo índice no es muy diferente de 1, de manera que E_a sería siempre mucho menor que E_s (para hacer nuestros cálculos más fáciles). Manteniendo nuestras aproximaciones deberíamos, por lo tanto, dejar de lado el término $\overline{E_a^2}$ porque es mucho menor que $\overline{E_sE_a}$. Ustedes podrían decir: "Entonces deberían dejar de lado $\overline{E_sE_a}$ también, porque es mucho menor que $\overline{E_s^2}$". Es cierto que $\overline{E_sE_a}$ es mucho menor que $\overline{E_s^2}$, pero debemos mantener $\overline{E_sE_a}$; ¡en caso contrario, nuestra aproximación sería la que se aplicaría si despreciáramos totalmente la presencia del material! Una manera de verificar que nuestros cálculos son compatibles es que siempre mantengamos términos que sean proporcionales a $N\Delta z$, la densidad de los átomos en el material, pero dejemos de lado términos que son proporcionales a $(N\Delta z)^2$ o cualquier otra potencia superior de $N\Delta z$. La nuestra es lo que debiera llamarse "aproximación de baja densidad".

En el mismo espíritu podríamos hacer notar que nuestra ecuación de energía ha despreciado la energía de la onda reflejada. Pero eso está bien, porque este término también es proporcional a $(N\Delta z)^2$, ya que la amplitud de la onda reflejada es proporcional a $N\Delta z$.

Para el último término en la ecuación (31.23) deseamos calcular la velocidad con que la onda incidente realiza trabajo sobre los electrones. Sabemos que trabajo es fuerza por distancia, de manera que la *velocidad* de realizar trabajo (también llamada potencia) es la fuerza por la velocidad. Realmente es $\boldsymbol{F} \cdot \boldsymbol{v}$, pero no necesitamos preocuparnos por el producto escalar, cuando la velocidad y la fuerza están en la misma dirección como lo están aquí (con excepción de un posible signo menos). Así que para cada átomo tomamos $\overline{q_eE_sv}$ como la velocidad promedio de realizar trabajo. Como existen $N\Delta z$ átomos por unidad de área, el último término en la ecuación (31.23) debe ser $N\Delta z q_e\overline{E_sv}$. Nuestra ecuación de energía es ahora:

$$\alpha\overline{E_s^2} = \alpha\overline{E_s^2} + 2\alpha\overline{E_sE_a} + N\,\Delta z\,q_e\overline{E_sv}. \tag{31.24}$$

Los términos $\overline{E_s^2}$ se simplifican y tenemos:

$$2\alpha\overline{E_sE_a} = -N\,\Delta z\,q_e\overline{E_sv}. \tag{31.25}$$

Ahora volvemos a la ecuación (30.19), que nos dice que para z grande

$$E_a = -\frac{N\,\Delta z\,q_e}{2\epsilon_0 c}\,v(\text{retard. en } z/c) \tag{31.26}$$

(recordando que $\eta = N\Delta z$). Introduciendo la ecuación (31.26) en el primer miembro de (31.25) obtenemos

$$-2\alpha \, \frac{N \, \Delta z \, q_e}{2\epsilon_0 c} \, \overline{E_s(\text{en } z) \cdot v(\text{retard. en } z/c)}.$$

Sin embargo, E_s (en z) es E_s (en los átomos) retardado en z/c. Ya que el promedio es independiente del tiempo, será igual ahora que retardado en z/c, o sea $\overline{E_s(\text{en el átomo}) \cdot v}$, el mismo promedio que aparece en el segundo miembro de (31.25). Los miembros son por lo tanto iguales, si

$$\frac{\alpha}{\epsilon_0 c} = 1 \qquad \text{o} \qquad \alpha = \epsilon_0 c. \tag{31.27}$$

Hemos descubierto que si la energía ha de conservarse, la energía transportada en una onda eléctrica por unidad de área y por unidad de tiempo (o sea, lo que hemos llamado *intensidad*) debe estar dada por $\varepsilon_0 c\overline{E^2}$. Si llamamos $\overline{S}$ a la intensidad, tenemos:

$$\overline{S} = \left\{ \begin{matrix} \text{intensidad} \\ \text{o} \\ \text{energía/área/tiempo} \end{matrix} \right\} = \epsilon_0 c \overline{E^2}, \tag{31.28}$$

donde la *barra* significa *promedio temporal*. ¡Tenemos un bonito resultado adicional de nuestra teoría del índice de refracción!

31-6 Difracción de la luz por una pantalla

Ahora es un buen momento para tratar un tema ligeramente diferente que podemos manejar con el mecanismo de este capítulo. En el último capítulo dijimos que cuando se tiene una pantalla opaca y la luz puede pasar a través de algunos agujeros, la distribución de la intensidad —el diagrama de difracción— podía obtenerse imaginando en su lugar que los agujeros eran remplazados por fuentes (osciladores) uniformemente distribuidas sobre el agujero. En otras palabras, la onda difractada es la misma que si el agujero fuera una nueva fuente. Tenemos que explicar la razón de eso, porque el agujero está, por supuesto, precisamente donde *no* hay fuentes, donde *no* hay cargas que aceleren.

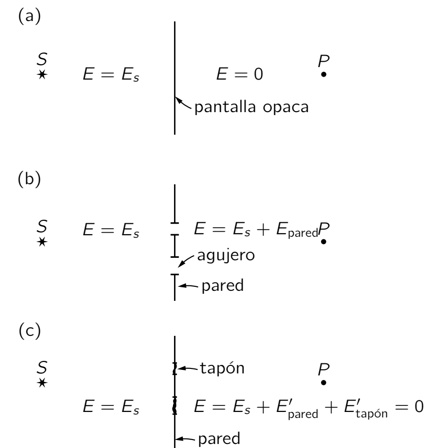


Figura 31-6. Difracción por una pantalla.

Preguntemos primero: "¿Qué *es* una pantalla opaca?" Supongan que tenemos una pantalla completamente opaca entre una fuente S y un observador en P, como en la figura 31-6(a). Si la pantalla es "opaca" no hay campo en P. ¿Por qué no hay campo ahí? De acuerdo con los principios básicos deberíamos obtener el campo en P como el campo retrasado E_s de la fuente más el campo de todas las otras cargas alrededor. Pero como hemos visto antes, las cargas de la pantalla serán puestas en movimiento por el campo E_s y estos movimientos generan un nuevo campo que, si la pantalla es opaca, debe *anular exactamente* el campo E_s en la parte trasera de la pantalla. Ustedes dirán: "¡Qué milagro que se anulen *exactamente*! ¡Supóngase que no fuera exactamente cierto!" Si no fuera exactamente cierto (recuerden que esta pantalla opaca tiene algún espesor), el campo hacia la parte de atrás de la pantalla no sería exactamente cero. Por consiguiente, no siendo cero, pondría en movimiento algunas otras cargas en el material de la pantalla y produciría así un poco más de campo, tratando de romper el balance. Por lo tanto, si hacemos la pantalla lo suficientemente gruesa no hay campo residual, porque hay suficiente oportunidad para conseguir que las cosas se aquieten finalmente. En términos de nuestras fórmulas anteriores diríamos que la pantalla tiene un índice imaginario grande, de manera que la onda es absorbida exponencialmente a medida que atraviesa. Ustedes saben, por supuesto, que una lámina suficientemente delgada del material más opaco, aun oro, es transparente.

Veamos ahora qué pasa con una pantalla opaca que tiene agujeros como en la figura 31-6(b) ¿Qué esperamos para el campo en P? El campo en P se puede representar como la suma de dos partes —el campo debido a la fuente S más el

campo debido a la pared—, es decir, debido al movimiento de las cargas en las paredes. Podemos esperar que el movimiento de las cargas en las paredes sea complicado, pero podemos averiguar *qué campos producen esto* de una manera bastante simple.

Supongan que tomáramos la misma pantalla, pero le taponamos los agujeros como se indica en la parte (c) de la figura. Imaginamos que los tapones son exactamente del mismo material que la pared. Noten que los tapones van donde estaban los agujeros en el caso (b). Calculemos ahora el campo en P. El campo en P es ciertamente cero en el caso (c), pero *también* es igual al campo de la fuente más el campo debido a todos los movimientos de los átomos en las paredes y en los tapones. Podemos escribir las ecuaciones siguientes:

$$\begin{aligned}
&\textit{Caso}\ \text{(b)}: \quad E_{\text{en }P} = E_s + E_{\text{pared}},\\
&\textit{Caso}\ \text{(c)}: \quad E'_{\text{en }P} = 0 = E_s + E'_{\text{pared}} + E'_{\text{tapón}},
\end{aligned}$$

donde las primas se refieren al caso cuando los tapones están en su lugar, pero E_s, por supuesto, es el mismo en ambos casos. Si restamos ambas ecuaciones obtenemos:

$$E_{\text{en }P} = (E_{\text{pared}} - E'_{\text{pared}}) - E'_{\text{tapón}}.$$

Ahora bien, si los agujeros no son demasiado chicos (digamos varias longitudes de onda transversalmente), no deberíamos esperar que la presencia de los tapones cambiara los campos que llegan a las paredes, excepto posiblemente en un pequeño monto alrededor de los bordes de los agujeros. Despreciando este pequeño efecto, podemos poner $E_{\text{pared}} = E'_{\text{pared}}$ y obtenemos que

$$E_{\text{en }P} = -E'_{\text{tapón}}.$$

¡Tenemos el resultado que el campo en P *cuando existen agujeros* en una pantalla (caso b) es el mismo (con excepción del signo) que el campo producido por *la parte* de una pared completamente opaca que está *localizada donde están los agujeros*! (El signo no es muy interesante ya que generalmente estamos interesados en la intensidad que es proporcional al cuadrado del campo). Esto parece una sorprendente argumentación en reversa. Sin embargo, no es únicamente válida (aproximadamente para agujeros no demasiado pequeños), sino útil y es la justificación de la teoría general de difracción.

El campo $E'_{\text{tapón}}$ se calcula en cualquier caso particular, recordando que el movimiento de las cargas en *todas* partes en la pantalla es aquel que va a anular el campo E_s en la parte de atrás de la pantalla. Una vez que conocemos estos movimientos, sumamos los campos de radiación en P debido sólo a las cargas en los tapones.

Hacemos notar nuevamente que esta teoría de la difracción es sólo aproximada y será válida sólo si los agujeros no son demasiado pequeños. Para agujeros que son demasiado pequeños, el término $E'_{\text{tapón}}$ será pequeño y entonces la diferencia entre E'_{pared} y E_{pared} (cuya diferencia hemos tomado como cero) puede ser comparable o mayor que el pequeño término $E'_{\text{tapón}}$ y nuestra aproximación ya no será válida.

32

Amortiguamiento por radiación. Dispersión de la luz

32-1 Resistencia de radiación

En el último capítulo aprendimos que cuando un sistema oscila expulsa energía, y dedujimos una fórmula para la energía irradiada por un sistema oscilante. Si conocemos el campo eléctrico, el promedio del cuadrado del campo multiplicado por $\epsilon_0 c$ es la cantidad de energía que pasa por metro cuadrado por segundo a través de una superficie normal a la dirección en que va la radiación:

$$S = \epsilon_0 c\langle E^2\rangle. \tag{32.1}$$

Cualquier carga oscilante irradia energía; por ejemplo, una antena excitada irradia energía. Si el sistema irradia energía, entonces, a fin de dar cuenta de la conservación de la energía, debemos encontrar que se esté dando potencia a través de los alambres que conducen a la antena. O sea, para el circuito de alimentación, la antena actúa como una *resistencia*, o un lugar donde la energía puede "perderse" (la energía no se pierde realmente, se irradia hacia afuera. Pero en lo que concierne al circuito, la energía se pierde). En una resistencia ordinaria, la energía que se "pierde" se transforma en calor; en este caso la energía que se "pierde" sale al espacio. Pero desde el punto de vista de la teoría de circuitos, sin considerar *a dónde* va la energía, el efecto neto sobre el circuito es el mismo —se "pierde" energía de ese circuito—. Por lo tanto, para el generador, la antena parece como si tuviera una resistencia, aun cuando puede estar hecha de cobre perfectamente bueno. En realidad si está bien construida, será casi como una resistencia pura, con muy poca inductancia o capacitancia, porque nos gustaría radiar la mayor cantidad de energía posible con la antena. Esta resistencia que exhibe la antena se llama *resistencia de radiación*.

Si a la antena llega una corriente I, la rapidez media con que se suministra potencia a la antena es el promedio del cuadrado de la corriente multiplicado por la resistencia. La rapidez con que la antena *irradia* potencia es proporcional al cuadrado de la corriente en la antena, porque todos los campos son proporcionales a las corrientes, y la energía liberada es proporcional al cuadrado del campo. El coeficiente de proporcionalidad entre la potencia irradiada e $\langle I^2\rangle$ es la resistencia de radiación.

Una pregunta interesante es: ¿a qué se debe esta resistencia de radiación? Tomemos un ejemplo sencillo: digamos que las corrientes están forzadas de un extremo al otro en una antena. Encontramos que debemos entregar trabajo para que la antena irradie energía. Si tomamos un cuerpo cargado y lo aceleramos hacia arriba y hacia abajo, irradia energía, y si no estuviera cargado, no irradiaría energía. Una cosa es calcular a partir de la conservación de energía que se pierde energía, pero otra cosa es contestar la pregunta: *¿contra qué fuerza* estamos haciendo trabajo? Ésa es una pregunta interesante y muy difícil que nunca ha sido completa y satisfactoriamente contestada para los electrones, aunque lo ha sido para las antenas. Lo que sucede es esto: en una antena los campos producidos por las cargas que se muevan en una parte de la antena reaccionan sobre las cargas que se muevan en otra parte de la antena. Podemos calcular esas fuerzas y hallar cuánto trabajo hacen, y así encontramos la regla correcta para la resistencia de radiación. Cuando decimos "podemos calcular", eso no es correcto, *nosotros* no podemos, porque aún no hemos estudiado las leyes de electricidad para distancias cortas. Solamente para distancias grandes sabemos cuánto vale

el campo eléctrico. Vimos la fórmula (28.3), pero actualmente es demasiado complicado para *nosotros* calcular los campos dentro de la zona de ondas. Por cierto, puesto que la conservación de la energía es válida, podemos calcular el resultado perfectamente sin conocer los campos a distancias cortas. En realidad, usando este razonamiento hacia atrás, resulta que uno puede encontrar la fórmula para las fuerzas a cortas distancias, sólo conociendo el campo a distancias muy largas, usando las leyes de conservación de la energía, pero no iremos a eso aquí.

En el caso de un simple electrón, el problema es éste: si hay solamente una carga, ¿sobre qué puede actuar la fuerza? Ha sido propuesto, en la vieja teoría clásica, que la carga era una pequeña bola y que una parte de la carga actuaba sobre la otra parte. Debido al retardo de la acción a través del pequeñísimo electrón, la fuerza no está exactamente en fase con el movimiento. Esto es, si tenemos el electrón quieto, sabemos que la "acción es igual a la reacción". Así las varias fuerzas internas son iguales y no hay una fuerza resultante. Pero si el electrón está acelerando, hay un retardo de tiempo a través de él y, por lo tanto, la fuerza que está actuando sobre el frente desde atrás no es la misma que la fuerza sobre la parte de atrás desde el frente, debido al retardo en el efecto. Este retardo en el tiempo es responsable de una falta de equilibrio, de manera que, como efecto neto, ¡la cosa es como si uno quisiera levantarse tirando de los cordones de los zapatos! Este modelo del origen de la resistencia a la aceleración, la resistencia de radiación de una carga móvil, ha encontrado muchas dificultades, porque nuestro punto de vista actual es que el electrón *no* es una "pequeña bola", este problema nunca ha sido resuelto. Sin embargo, podemos calcular exactamente, por supuesto, cuál debe ser la resistencia de radiación resultante, es decir, cuánta pérdida debe haber cuando aceleramos una carga a pesar de no conocer directamente el mecanismo de cómo funciona esa fuerza.

32-2 La rapidez de radiación de energía

Ahora calcularemos la energía total irradiada por una carga acelerada. Para mantener la generalidad de la discusión tomaremos el caso de una carga acelerando de cualquier forma, pero no en forma relativista. En el momento en que la aceleración es, digamos, vertical, sabemos que el campo eléctrico que se genera es la carga multiplicada por la proyección de la aceleración retardada, dividida entre la distancia. Así, pues, conocemos el campo eléctrico en cualquier punto y, por lo tanto, conocemos el cuadrado del campo eléctrico y, en consecuencia, la energía $\epsilon_0 c E^2$ que escapa por unidad de área por segundo.

La cantidad $\epsilon_0 c$ aparece bastante frecuentemente en expresiones que tienen que ver con propagación de ondas de radio. Su inversa se llama *impedancia del vacíos*, y es un número fácil de recordar: tiene el valor $1/\epsilon_0 c = 377$ ohms. Luego la potencia en watts por metro cuadrado es igual al promedio del cuadrado del campo dividido por 377.

Usando nuestra expresión (29.1) para el campo eléctrico, encontramos que

$$S = \frac{q^2 a'^2 \operatorname{sen}^2 \theta}{16\pi^2 \epsilon_0 r^2 c^3} \tag{32.2}$$

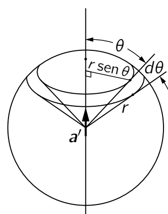


Figura 32-1. El área de un sector esférico es $2\pi r \operatorname{sen}\theta \cdot r\,d\theta$.

es la potencia por metro cuadrado irradiada en la dirección θ. Notamos que varía inversamente con el cuadrado de la distancia, como dijimos antes. Ahora supongamos que queremos la energía total irradiada en todas direcciones; entonces debemos integrar la ecuación (32.2) sobre todas las direcciones. Primero multiplicamos por el área, para encontrar la cantidad que fluye dentro de un pequeño ángulo $d\theta$ (véase figura 32-1). Necesitamos el área de una sección esférica. La manera de obtenerla es ésta: si r es el radio, el ancho del segmento anular es $r\,d\theta$ y la circunferencia es $2\pi r \operatorname{sen}\theta$, porque $r \operatorname{sen}\theta$ es el radio del círculo. Así, pues, el área del pequeño pedazo de esfera es $2\pi r \operatorname{sen}\theta$ multiplicado por $rd\theta$:

$$dA = 2\pi r^2 \operatorname{sen}\theta\,d\theta. \tag{32.3}$$

Multiplicando el flujo [ecuación (32.2), la potencia por metro cuadrado] por el área en metros cuadrados incluida en el pequeño ángulo $d\theta$, encontramos

la cantidad de energía que se libera en esta dirección entre θ y $\theta + d\theta$; luego integramos sobre todos los ángulos θ de 0 a 180°:

$$P = \int S\,dA = \frac{q^2 a'^2}{8\pi\epsilon_0 c^3} \int_0^\pi \operatorname{sen}^3\theta\,d\theta. \tag{32.4}$$

Escribiendo $\operatorname{sen}^3\theta = (1 - \cos^2\theta)\operatorname{sen}\theta$, no es difícil ver que $\int_0^\pi \operatorname{sen}^3\theta d\theta = 4/3$. Usando esto obtenemos, finalmente,

$$P = \frac{q^2 a'^2}{6\pi\epsilon_0 c^3}. \tag{32.5}$$

Esta expresión merece algunas observaciones. Ante todo, puesto que el vector $\boldsymbol{a}'$ tenía cierta dirección, el a'^2 en (32.5) debería ser el cuadrado del vector $\boldsymbol{a}'$, esto es, $\boldsymbol{a}' \cdot \boldsymbol{a}'$, la longitud del vector al cuadrado. Segundo, el flujo (32.2) fue calculado usando la aceleración retardada, esto es, la aceleración en el tiempo en el cual la energía que ahora está pasando a través de la esfera fue irradiada. Nos gustaría decir que esta energía fue en realidad liberada en este tiempo anterior. Esto no es totalmente cierto, es sólo una idea aproximada. El momento exacto en que la energía es liberada no se puede definir precisamente. Todo lo que podemos realmente calcular en forma precisa es qué sucede en un movimiento completo, como una oscilación o algo donde la aceleración finalmente cesa. Entonces, lo que encontramos es que el flujo total de energía por ciclo es el promedio de la aceleración al cuadrado para un ciclo completo. Esto es lo que debería realmente aparecer en (32.5). O bien, si es un movimiento con una aceleración que es inicial y finalmente cero, la energía total que ha escapado es la integral de (32.5) sobre el tiempo.

Para ilustrar las consecuencias de la fórmula (32.5) cuando tenemos un sistema oscilante, veamos qué sucede si el desplazamiento x de la carga está oscilando, de manera que la aceleración a es $-\omega^2 x_0 e^{i\omega t}$. El promedio de la aceleración al cuadrado sobre un ciclo (recuerden que tenemos que ser muy cuidadosos cuando elevamos al cuadrado cosas que están escritas en una notación compleja —realmente es el coseno—, y el promedio de $\cos^2\omega t$ es un medio) es entonces

$$\langle a'^2 \rangle = \tfrac{1}{2}\omega^4 x_0^2.$$

Por tanto

$$P = \frac{q^2 \omega^4 x_0^2}{12\pi\epsilon_0 c^3}. \tag{32.6}$$

Las fórmulas que estamos discutiendo son relativamente avanzadas y más o menos modernas; datan del comienzo del siglo XX, y son muy famosas. Debido a su valor histórico, es importante que nosotros podamos leer acerca de ellas en libros más antiguos. En realidad, los libros antiguos también usaban un sistema de unidades diferente de nuestro sistema mks actual. Sin embargo, todas esas complicaciones pueden ser arregladas en las fórmulas finales que tratan de electrones mediante la siguiente regla: la cantidad $q_e^2/4\pi\epsilon_0$, donde q_e es la carga electrónica (en coulombs), ha sido históricamente escrita como e^2. Es muy fácil calcular que e en el sistema mks es numéricamente igual a 1.5188×10^{-14}, porque sabemos que numéricamente $q_e = 1.60206 \times 10^{-19}$ y $1/4\pi\epsilon_0 = 8.98748 \times 10^9$. Por lo tanto, a menudo usaremos la abreviatura conveniente:

$$e^2 = \frac{q_e^2}{4\pi\epsilon_0}. \tag{32.7}$$

Si usamos el valor numérico anterior en las fórmulas antiguas y las tratamos como si estuvieran escritas en unidades mks, obtendremos los resultados numéricos correctos. Por ejemplo, la forma antigua de (32.5) es $P = \frac{2}{3}e^2 a'^2/c^3$. De nuevo la energía potencial de un protón y un electrón a una distancia r es $q_e^2/4\pi\epsilon_0 r$ o e^2/r con $e = 1.5188 \times 10^{-14}$ (mks).

32-3 Amortiguamiento de la radiación

Ahora bien, el hecho de que un oscilador pierda cierta energía debería significar que si tuviéramos una carga en el extremo de un resorte (o un electrón en un átomo), el cual tiene una frecuencia natural ω_0, y comenzáramos a oscilarlo y lo soltáramos, no oscilaría indefinidamente, aun si estuviera en un espacio vacío a millones de kilómetros de cualquier cosa. No hay aceite, ninguna resistencia, en un sentido ordinario; no hay "viscosidad". Pero, sin embargo, no oscilará, como podríamos haber dicho en una ocasión, "para siempre", porque si se encuentra cargado, está irradiando energía, y por lo tanto la oscilación lentamente se detendrá. ¿Con qué lentitud? ¿Cuál es el Q de tal oscilador, causado por los efectos electromagnéticos, la así llamada resistencia de radiación o amortiguamiento por radiación de un oscilador? El Q de cualquier sistema oscilante es el total del contenido de energía de un oscilador en cualquier momento dividido entre la energía perdida por radián:

$$Q = -\frac{W}{dW/d\phi}.$$

O bien (otra forma de escribirlo), puesto que $dW/d\phi = (dW/dt)/(d\phi/dt) = (dW/dt)/\omega$,

$$Q = -\frac{\omega W}{dW/dt}. \tag{32.8}$$

Si para un Q dado esto nos dice cómo se extingue la energía de la oscilación, $dW/dt = -(\omega/Q)W$, que tiene la solución $W = W_0 e^{-\omega t/Q}$ si W_0 es la energía inicial (para $t = 0$).

Para encontrar el Q de un radiador, volvemos a (32.8) y usamos (32.6) para dW/dt.

Ahora bien, ¿qué usamos para la energía W del oscilador? La energía cinética del oscilador es $\frac{1}{2}mv^2$, y la energía cinética *media* es $m\omega^2 x_0^2/4$. Pero recordemos que la energía total de un oscilador es, en promedio, mitad energía cinética y mitad energía potencial, por lo que duplicamos nuestro resultado y encontramos para la energía total del oscilador

$$W = \tfrac{1}{2}m\omega^2 x_0^2. \tag{32.9}$$

¿Qué usamos para la frecuencia en nuestras fórmulas? Usamos la frecuencia natural w_0, porque, para todo fin práctico, ésa es la frecuencia a la cual nuestro átomo está irradiando, y para m usamos la masa m_e del electrón. Luego, haciendo las divisiones y cancelaciones necesarias, la fórmula se reduce a

$$\frac{1}{Q} = \frac{4\pi e^2}{3\lambda m_e c^2}. \tag{32.10}$$

(Para verlo mejor y en una forma más histórica, la escribimos usando nuestra abreviatura $q_e^2/4\pi\epsilon_0 = e^2$ y el factor ω_0/c que sobró ha sido escrito como $2\pi/\lambda$.) Como Q no tiene dimensión, la combinación $e^2/m_e c^2$ debe ser una propiedad solamente de la carga y de la masa del electrón, una propiedad intrínseca del electrón, y debe ser una *longitud*. Se le ha dado un nombre, el *redio clásico del electrón*, porque los modelos atómicos primitivos, que fueron inventados para explicar la resistencia de radiación sobre la base de la fuerza de una parte del electrón actuando sobre las otras partes, necesitaban todos tener un electrón cuyas dimensiones eran de este orden general de magnitud. Sin embargo, esta cantidad ya no tiene el significado de que creemos que el electrón realmente *tiene* tal radio. Numéricamente, la magnitud del radio es

$$r_0 = \frac{e^2}{m_e c^2} = 2.82 \times 10^{-15} \text{ m}. \tag{32.11}$$

Calculemos ahora realmente el Q de un átomo que está emitiendo luz —digamos un átomo de sodio—. Para un átomo de sodio, la longitud de onda es

aproximadamente 6 000 ángstroms, en la parte amarilla del espectro visible, y es una longitud de onda típica. Entonces

$$Q = \frac{3\lambda}{4\pi r_0} \approx 5 \times 10^7, \tag{32.12}$$

así el Q de un átomo es del orden de 10^8. Esto significa que un oscilador atómico oscilará por 10^8 radianes, o cerca de 10^7 oscilaciones, antes de que su energía decrezca en un factor de $1/e$. La frecuencia de oscilación de luz correspondiente a 6 000 ángstroms, $v = c/\lambda$, es del orden de 10^{15} ciclos/segundo, y por lo tanto la vida media, el tiempo que demora la energía de un átomo radiante en disminuir en un factor de $1/e$, es del orden de 10^{-8} segundos. En circunstancias ordinarias, los átomos que emiten libremente demoran por lo general este tiempo en irradiar. Esto es válido solamente para átomos que están en el espacio vacío, no estando perturbados en forma alguna. Si el electrón está en un sólido y tiene que golpear a otros átomos o electrones, entonces hay resistencias adicionales y diferente amortiguamiento.

El término resistencia efectiva γ en la ley de resistencia de un oscilador puede encontrarse usando la relación $1/Q = \gamma/\omega_0$, y recordamos que el valor de γ determina lo ancha que es la curva de resonancia (figura 23-2). ¡Así, pues, hemos calculado precisamente los *anchos de las líneas espectrales* para átomos que irradian libremente! Como $\lambda = 2\pi c/\omega$, encontramos que

$$\begin{aligned}\Delta\lambda &= 2\pi c\,\Delta\omega/\omega^2 = 2\pi c\gamma/\omega_0^2 = 2\pi c/Q\omega_0\\ &= \lambda/Q = 4\pi r_0/3 = 1.18 \times 10^{-14}\text{ m.}\end{aligned} \tag{32.13}$$

32-4 Fuentes independientes

En preparación para nuestro segundo tópico, la dispersión de la luz, debemos analizar ahora un cierto aspecto del fenómeno de interferencia que omitimos discutir anteriormente. Ésta es la interrogante de cuándo la interferencia *no* tiene lugar. Si tenemos dos fuentes S_1 y S_2, con amplitudes A_1 y A_2, y hacemos una observación en una cierta dirección en la cual las fases de llegada de las dos señales son ϕ_1 y ϕ_2 (una combinación del tiempo real de oscilación y el tiempo retardado, dependiendo de la posición de observación), entonces la energía que recibimos se puede encontrar componiendo los dos vectores o números complejos A_1 y A_2, uno con ángulo ϕ_1 y el otro con ángulo ϕ_2 (como hicimos en el capítulo 29) y encontramos que la energía resultante es proporcional a

$$A_R^2 = A_1^2 + A_2^2 + 2A_1A_2\cos(\phi_1 - \phi_2). \tag{32.14}$$

Ahora bien, si el término cruzado $2A_1A_2\cos(\phi_1 - \phi_2)$ no estuviera, la energía total que sería recibida en una dirección dada sería simplemente la suma de las energías, $A_1^2 + A_2^2$, que sería liberada por cada fuente separadamente, que es lo que normalmente esperamos. Esto es, la intensidad combinada de luz brillando sobre algo proveniente de dos fuentes es la suma de las intensidades de las dos luces. Por otro lado, si tenemos las cosas en forma correcta y tenemos un término cruzado, ¿no es tal suma? porque hay también alguna interferencia. Si hay circunstancias en las cuales este término no es de importancia, diremos que aparentemente se ha perdido la interferencia. Por cierto, en esencia está siempre ahí, pero puede ser que no seamos capaces de detectarla.

Consideremos algunos ejemplos. Supongamos, primero, que las dos fuentes están 7 000 millones de longitudes de onda una de otra, una disposición no imposible. Entonces en una dirección dada es cierto que hay un valor muy definido de estos desfases. Pero, en cambio, si nos movemos solamente un pelo en una dirección, unas pocas longitudes de onda, lo cual no es distancia en absoluto (nuestro ojo ya tiene un orificio que es tan grande que estamos promediando los efectos sobre un intervalo muy amplio comparado con una longitud de onda), entonces cambiamos la fase relativa, y el coseno cambia muy rápidamente. Si tomamos el *promedio* de la intensidad sobre una región pequeña, entonces el

coseno, que toma valores más, menos, más, menos, cuando nos desplazamos, da un promedio nulo.

Luego, si promediamos sobre regiones donde la fase varía muy rápidamente con la posición, no obtenemos interferencia.

Otro ejemplo. Supongamos que las dos fuentes son dos osciladores de radio independientes —no un solo oscilador alimentado por dos alambres, que garantiza que las fases se mantienen juntas, sino dos fuentes independientes— y que no están *exactamente* sintonizados a la misma frecuencia; es muy difícil hacerlas coincidir exactamente en la misma frecuencia sin interconectarlos realmente. En este caso tenemos lo que llamamos dos fuentes *independientes*. Naturalmente, como las frecuencias no son exactamente iguales, aunque partieron en fase, una de ellas comienza a estar un poco delante de la otra, y muy pronto ellas están desfasadas, y luego adelanta aún más, y muy pronto están en fase de nuevo. Así, el desfase entre las dos está gradualmente desplazándose con el tiempo, pero si nuestra observación es tan tosca que no podemos observar aquel pequeño tiempo, si promediamos sobre un tiempo mucho más largo, entonces aunque la intensidad aumente y disminuya como lo que llamamos "pulsaciones" en sonido, si estos aumentos y disminuciones son demasiado rápidos para que los sigan nuestros equipos, entonces de nuevo este término se promedia eliminándose.

En otras palabras, ¡en cualquier circunstancia en que el desfase se promedia a cero, no obtenemos interferencia!

Se encuentran muchos libros que dicen que dos fuentes de luz diferentes nunca interfieren. Esto no es un enunciado físico, sino meramente un enunciado sobre el grado de sensibilidad de la técnica de los experimentos en la época en que se escribió el libro. Lo que sucede en una fuente luminosa es que primero un átomo irradia, luego otro átomo irradia, y así sucesivamente, y ya hemos visto que los átomos irradian un tren de ondas en solamente alrededor de 10^{-8} s; después de 10^{-8} s, algún átomo probablemente ha entrado en juego, luego otro átomo, y así sucesivamente. Así, pues, las fases pueden en realidad permanecer iguales sólo por aproximadamente 10^{-8} s. Por lo tanto, si promediamos por mucho más de 10^{-8} s, no vemos una interferencia de dos fuentes diferentes, porque no pueden mantener sus fases fijas por más de 10^{-8} s. Con fotocélulas es posible la detección de muy alta rapidez, y se puede demostrar que hay una interferencia que varía con el tiempo, aumentando y disminuyendo en alrededor de 10^{-8} s. Pero la mayoría de los equipos de detección, desde luego, no observan intervalos de tiempo tan finos y, por tanto, no ven interferencia. Ciertamente con el ojo, que promedia sobre un tiempo de un décimo de segundo, no hay posibilidad alguna de ver interferencia en dos fuentes ordinarias diferentes.

Recientemente ha sido posible hacer fuentes luminosas que evitan este efecto haciendo que todos los átomos emitan *juntos* en el tiempo. El dispositivo que hace esto es una cosa muy complicada, y debe ser comprendida en una manera cuántica. Se llama *láser*, y es posible producir a partir de un láser una fuente en la cual el tiempo en el que la fase se mantiene constante es mucho más larga que 10^{-8} s. Puede ser del orden de un centésimo, un décimo, o aun de un segundo, y así, con fotocélulas ordinarias, se puede detectar la frecuencia entre dos diferentes láseres. Se puede detectar fácilmente el latido de las pulsaciones entre dos fuentes láser. ¡Pronto, sin duda, alguien será capaz de mostrar dos fuentes iluminando una pared, en las cuales las pulsaciones son tan lentas que uno pueda ver que la pared se ilumina y se oscurece!

Otro caso en el cual la interferencia promedio se anula es aquel en el cual, en vez de tener solamente *dos* fuentes, tenemos *muchas*. En este caso, escribiríamos la expresión para A_R^2 como la suma de muchas amplitudes, números complejos, elevada al cuadrado, y obtendríamos el cuadrado de cada uno, todos sumados, más términos cruzados entre cada par, y si las circunstancias son tales que los últimos se promedian a cero, entonces no habrá efectos de interferencia. Puede ser que las diversas fuentes estén ubicadas en posiciones tan al azar que, aunque el desfase entre A_2 y A_3 sea también definido, es muy diferente del que hay entre A_1 y A_2, etc. Obtendríamos así muchos cosenos, muchos positivos, muchos negativos, todos promediando a cero.

Así es que en muchas circunstancias no vemos los efectos de interferencia, sino que vemos sólo una intensidad total colectiva, igual a la suma de todas las intensidades.

32-5 Dispersión de la luz

Lo anterior nos conduce a un efecto que ocurre en el aire como consecuencia de las posiciones irregulares de los átomos. Cuando estábamos discutiendo el índice de refracción, vimos que un haz de luz que llega hará que los átomos irradien de nuevo. El campo eléctrico del haz que llega mueve los electrones de un lado al otro, e irradian debido a sus aceleraciones. Esta radiación dispersada se combina para dar un haz en la misma dirección del haz original, pero de una fase algo diferente, y éste es el origen del índice de refracción.

¿Pero qué podemos decir acerca de la cantidad de luz que vuelve a irradiar en alguna otra dirección? Normalmente, si los átomos están muy bien colocados en una buena red, es fácil demostrar que no obtendremos cosa alguna en otra dirección porque estamos sumando muchos vectores con sus fases siempre cambiantes, y el resultado se hace cero. Pero si los objetos están ubicados *al azar*, la intensidad total en cualquier dirección es la *suma* de intensidades dispersadas por cada átomo, como ya hemos discutido. Además, los átomos de un gas están realmente en movimiento, así que aunque la fase relativa de los átomos sea ahora una cantidad definida, más tarde será bastante diferente y, por lo tanto, *cada* término coseno se promediará a cero. Por lo tanto, para encontrar cuánta luz es dispersada por un gas en una dirección dada, simplemente estudiamos los efectos de *un átomo* y multiplicamos la intensidad que éste irradia por el número de átomos.

Anteriormente señalamos que el fenómeno de dispersión de la luz de esta naturaleza es el origen del azul del cielo. La luz del Sol pasa a través del aire, y cuando miramos hacia un lado del Sol —digamos a 90° con respecto al haz— vemos luz azul; lo que tenemos ahora que calcular es *cuánta* luz vemos y *por qué es azul.*

Si el haz incidente tiene el campo eléctrico* $\boldsymbol{E} = \hat{\boldsymbol{E}}_o e^{i\omega t}$ en el punto donde está ubicado el átomo, sabemos que un electrón en el átomo vibrará de un lado a otro en respuesta a este $\boldsymbol{E}$ (figura 32-2). Conforme a la ecuación (23.8), la respuesta será

$$\hat{\boldsymbol{x}} = \frac{q_e\hat{\boldsymbol{E}}_0}{m(\omega_0^2 - \omega^2 + i\gamma\omega)}. \tag{32.15}$$

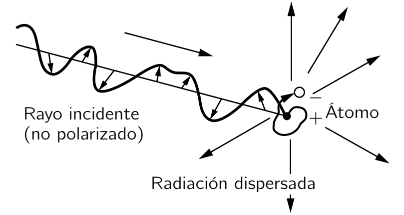


Figura 32-2. Un haz de radiación incide sobre un átomo y hace que las cargas (electrones) del átomo se muevan. Los electrones en movimiento a su vez irradian en varias direcciones.

Podríamos incluir el amortiguamiento y la posibilidad de que el átomo actúe como varios osciladores de frecuencias diferentes y sumar sobre las diversas frecuencias, pero por simplicidad tomemos simplemente un oscilador y despreciemos el amortiguamiento. Entonces la respuesta al campo eléctrico externo, la cual hemos ya usado en el cálculo del índice de refracción, es simplemente

$$\hat{\boldsymbol{x}} = \frac{q_e\hat{\boldsymbol{E}}_0}{m(\omega_0^2 - \omega^2)}. \tag{32.16}$$

Ahora podríamos calcular fácilmente la intensidad de luz emitida en varias direcciones, usando la fórmula (32.2) y la aceleración correspondiente al $\hat{\boldsymbol{x}}$ anterior.

Mejor calculemos sólo la *cantidad total* de luz dispersada en *todas* direcciones, a fin de ahorrar tiempo. La cantidad total de energía luminosa por segundo, dispersada en todas direcciones por un átomo simple, está naturalmente dada por la ecuación (32.6). Así, reuniendo las partes y reagrupándolas, obtenemos

$$\begin{aligned} P &= [(q_e^2\omega^4/12\pi\epsilon_0c^3)q_e^2E_0^2/m_e^2(\omega^2 - \omega_0^2)^2] \\ &= (\tfrac{1}{2}\epsilon_0cE_0^2)(8\pi/3)(q_e^4/16\pi^2\epsilon_0^2m_e^2c^4)[\omega^4/(\omega^2 - \omega_0^2)^2] \\ &= (\tfrac{1}{2}\epsilon_0cE_0^2)(8\pi r_0^2/3)[\omega^4/(\omega^2 - \omega_0^2)^2] \end{aligned} \tag{32.17}$$

para la potencia total dispersada, irradiada en todas direcciones.

* Cuando aparece un símbolo de intercalación en un vector, significa que las *componentes* del vector son complejos: $\hat{\boldsymbol{E}} = (\hat{E}_x, \hat{E}_y, \hat{E}_z)$.

Hemos escrito el resultado en la forma anterior porque entonces es fácil de recordar: primero, la energía total dispersada es proporcional al cuadrado del campo incidente. ¿Qué significa eso? Evidentemente, el cuadrado del campo incidente es proporcional a la energía que está llegando por segundo. En realidad, la energía incidente por metro cuadrado por segundo es $\epsilon_0 c$ veces el promedio $\langle E^2\rangle$ del cuadrado del campo eléctrico, y si E_0 es el valor máximo de E, $\langle E^2\rangle = \frac{1}{2}E_0^2$. En otras palabras, la energía total dispersada es proporcional a la energía por metro cuadrado que llega; cuanto más brillante sea la luz del Sol que ilumina el cielo, más brillante va a parecer el cielo.

Entonces, ¿qué *fracción* de la luz que llega es dispersada? Imaginemos un "blanco" con una cierta área, digamos σ, en el haz (no un blanco real o material, porque éste difractaría luz, etc., queremos decir un área imaginaria dibujada en el espacio). La cantidad total de energía que pasaría a través de esta superficie σ en una circunstancia dada es proporcional tanto a la intensidad que llega como a σ, y la potencia total sería

$$P = (\tfrac{1}{2}\epsilon_0 c E_0^2)\sigma. \tag{32.18}$$

Ahora inventamos una idea: decimos que el átomo dispersa una cantidad total de intensidad que es la cantidad que caería en una cierta área geométrica, y damos la respuesta dando esa área. Esa respuesta, entonces, es independiente de la intensidad incidente; da el cociente entre la energía dispersada y la energía incidente por metro cuadrado. En otras palabras, el cociente

$$\frac{\text{energía total dispersada por segundo}}{\text{energía incidente por metro cuadrado por segundo}}\ \text{es un } \textit{área}.$$

El significado de esta área es que, si toda la energía que incide sobre la misma fuera arrojada en todas direcciones, entonces ésa es la cantidad de energía que sería dispersada por el átomo.

Esta área se llama *sección transversal de dispersión*; el concepto de sección eficaz se usa constantemente, dondequiera que ocurra algún fenómeno proporcional a la intensidad de un haz. En tales casos, siempre se describe el monto del fenómeno diciendo cuál debería ser el área efectiva para recoger esa cantidad del haz. No significa de manera alguna que este oscilador *tenga* realmente tal área. Si no hubiera cosa alguna presente sino un electrón libre sacudiéndose de un lado al otro, no habría área directamente asociada con él físicamente. Es simplemente una forma de expresar la respuesta a cierto tipo de problema; nos dice cuánta área del haz incidente debería golpear a fin de dar cuenta de cuánta energía sale. Luego para nuestro caso

$$\sigma_s = \frac{8\pi r_0^2}{3}\cdot\frac{\omega^4}{(\omega^2-\omega_0^2)^2} \tag{32.19}$$

(el subíndice s indica "dispersión").

Veamos algunos ejemplos. Primero, si vamos a una frecuencia natural muy baja ω_0, o a electrones completamente no ligados, para los cuales $\omega_0 = 0$, la frecuencia ω es más simple y la sección eficaz es constante. Esta sección eficaz del electrón libre, o límite de baja frecuencia, se conoce como *sección eficaz de dispersión de Thomson*. ¡Es un área cuyas dimensiones son aproximadamente 10^{-15} metros, más o menos por lado, es decir, 10^{-30} metros cuadrados, ¡lo cual es bastante pequeño!

Por otra parte, si tomamos el caso de la luz en el aire, recordamos que para el aire las frecuencias naturales de los osciladores son mayores que la frecuencia de la luz que usamos. Esto significa que en una primera aproximación podemos hacer caso omiso de ω^2 en el denominador, y encontramos que la dispersión es proporcional a la *cuarta potencia* de la frecuencia. Es decir, la luz que es de frecuencia más alta en un factor de dos, digamos, es 16 *veces* más intensamente dispersada, lo que es una diferencia bastante considerable. Esto significa que la luz azul, que tiene alrededor del doble de la frecuencia del extremo rojizo del espectro, es dispersada en una proporción mucho mayor que la luz roja.

¡Entonces, cuando miramos al cielo, aparece aquel azul espléndido, que vemos todo el tiempo!

Hay varios puntos a señalar acerca de los resultados anteriores; una pregunta interesante es: ¿por qué llegamos a ver las *nubes*? ¿De dónde vienen las nubes? Todos saben que es la condensación del vapor de agua. Pero naturalmente el vapor de agua está ya en la atmósfera *antes* de que se condense; así que, ¿por qué no lo vemos entonces? Después que se condensa es perfectamente visible. No estaba ahí, ahora *está* ahí. Así que el misterio de dónde vienen las nubes no es realmente un misterio infantil como el de "¿de dónde viene el agua, papito?", sino que tiene que ser explicado.

Acabamos de explicar que todo átomo dispersa luz, y naturalmente el vapor de agua también dispersará luz. El misterio es: ¿por qué, cuando el agua se condensa en nubes, dispersa una *cantidad tremendamente grande* de luz?

Consideren qué sucedería si, en vez de un átomo simple, tuviéramos un aglomerado de átomos, digamos dos, muy cerca uno del otro comparado con la longitud de onda de la luz. Recuerden, los átomos son solamente de alrededor de un ángstrom, mientras que la longitud de onda de la luz es de unos 5 000 ángstroms, de modo que cuando forman un racimo, unos pocos átomos juntos, pueden *estar* muy cerca comparados con la longitud de onda de la luz. Entonces, cuando el campo eléctrico actúa, *ambos átomos se moverán juntos*. El campo eléctrico dispersado será entonces la suma de los dos campos eléctricos en fase, es decir, duplica la amplitud que había con un átomo simple, ¡y la *energía* que se dispersa es por lo tanto *cuatro veces* la correspondiente a un átomo simple, no dos veces! Así, pues, racimos de átomos irradian o dispersan más energía de lo que hacen cuando están como átomos simples. Nuestro argumento de que las fases son independientes está basado en la suposición que hay un desfase real y grande entre dos átomos cualesquiera, lo cual es cierto solamente si están a varias longitudes de onda uno de otro y espaciados al azar, o moviéndose. Pero si están próximos unos con otros, forzosamente dispersan en fase y tienen una interferencia coherente que produce un aumento en la dispersión.

Si tenemos N átomos en un racimo, lo cual es una pequeñísima gota de agua, cada uno será forzado por el campo eléctrico aproximadamente de la misma manera que antes (el efecto de un átomo sobre otro no es importante, es sólo para tener una idea) y la amplitud de dispersión de cada uno es la misma, de manera que el campo total dispersado aumenta N veces. La *intensidad* de la luz que se dispersa se incrementa entonces en el cuadrado, o sea N^2 veces. Habríamos esperado, si los átomos estuvieran dispersos en el espacio, solamente N veces 1, ¡mientras que tenemos N^2 veces 1! Es decir, la dispersión de agua en racimos de N moléculas cada uno es N veces más intensa que la dispersión de los átomos simples. En consecuencia, a medida que el agua se aglomera, la dispersión aumenta. ¿Se incrementa *indefinidamente*? ¡No! ¿Cuándo comienza este análisis a fallar? ¿Cuántos átomos podemos reunir antes de que no podamos llevar este razonamiento más adelante? *Respuesta*: si la gota de agua llega a ser tan grande que de un extremo al otro haya una longitud de onda o algo parecido, los átomos no estarán más en fase porque están demasiado separados. Así, a medida que seguimos aumentando el tamaño de las gotas, obtenemos más y más dispersión, hasta un instante tal en que una gota llega alrededor del tamaño de una longitud de onda, y entonces la dispersión no aumenta ni siquiera cercanamente tan rápido como aumenta la gota. Además, el azul desaparece, porque para longitudes de onda grandes las gotas pueden ser mayores, antes de que este límite sea alcanzado, de lo que pueden ser para longitudes de onda cortas. Aunque por átomo las ondas cortas se dispersan más que las ondas largas, hay un mayor realce para el extremo rojo del espectro que para el extremo azul cuando las gotas son más grandes que la longitud de onda, de manera que el color se corre del azul hacia el rojo.

Ahora podemos hacer un experimento que demuestra esto. Podemos hacer partículas que son muy pequeñas al comienzo, y luego crecen gradualmente en tamaño. Usamos una solución de tiosulfato de sodio (hipo) en ácido sulfúrico, que precipita granos muy finos de azufre. A medida que el azufre precipita, los granos

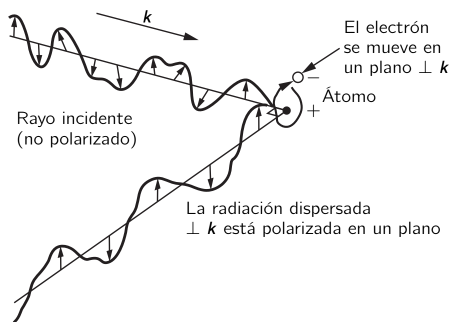


Figura 32-3. Ilustración del origen de la polarización de la radiación dispersada en ángulo recto con el rayo incidente.

primero comienzan muy finos, y la dispersión es un poco azulada. A medida que precipita más, se pone más intenso, y luego se pondrá blancuzco cuando las partículas se hacen más grandes. Además, la luz que atraviesa directamente tendrá eliminado el azul. Es por esto que la puesta del sol es roja, naturalmente, porque la luz que pasa a través de mucho aire para llegar al ojo tiene eliminada mucha luz azul por dispersión, por lo que es rojo-amarillenta.

Finalmente, hay otro aspecto importante que realmente pertenece al próximo capítulo sobre polarización, pero es tan interesante que lo destacamos ahora. Es que el campo eléctrico de la luz dispersada tiende a vibrar en una dirección particular. El campo eléctrico de la luz que llega está oscilando de alguna manera y el oscilador excitado va en esta misma dirección, y si estamos situados aproximadamente en ángulo recto con el haz, veremos luz *polarizada*, es decir, luz en la cual el campo eléctrico va solamente de una manera. En general, los átomos pueden vibrar en cualquier dirección perpendicular al rayo, pero si están forzados directamente hacia nosotros o alejándose, no los vemos. ¡Así, si la luz que llega tiene un campo eléctrico que cambia y oscila en cualquier dirección, la cual llamamos luz no polarizada, entonces la luz que sale a 90° respecto del rayo vibra solamente en una dirección! (véase la figura 32-3).

Existe una sustancia llamada polaroide que tiene la propiedad de que cuando la luz pasa por ella, solamente la parte del campo eléctrico que está a lo largo de un eje particular puede pasar. Podemos usar esto, para demostrar la polarización, y en verdad encontramos que la luz dispersada por la solución hipo está fuertemente polarizada.

33

Polarización

33-1 El vector eléctrico de la luz

En este capítulo consideraremos los fenómenos que dependen del hecho que el campo eléctrico que describe la luz es un vector. En capítulos anteriores no hemos estado interesados en la dirección de oscilación del campo eléctrico, excepto para notar que el vector eléctrico yace en un plano perpendicular a la dirección de propagación. La dirección particular en este plano no nos ha interesado. Consideramos ahora aquellos fenómenos cuyo aspecto central es la dirección particular de oscilación del campo eléctrico.

En la luz idealmente monocromática, el campo eléctrico debe oscilar a una frecuencia definida, pero como la componente x y la componente y pueden oscilar independientemente a una frecuencia definida, debemos considerar primero el efecto resultante producido al superponer dos oscilaciones independientes perpendiculares entre sí. ¿Qué clase de campo eléctrico está compuesto de un componente x y una componente y que oscila a la misma frecuencia? Si se suma a una vibración x una cierta cantidad de vibración y con la misma fase, el resultado es una nueva vibración en el plano xy. La figura 33-1 ilustra la superposición de amplitudes diferentes para la vibración x y la vibración y. Pero las resultantes mostradas en la figura 33-1 no son las únicas posibilidades; en todos estos casos hemos supuesto que la vibración x y la vibración y están *en fase*, pero no tiene por qué ser de esa manera. Podría ser que la vibración x y la vibración y estuvieran fuera de fase.

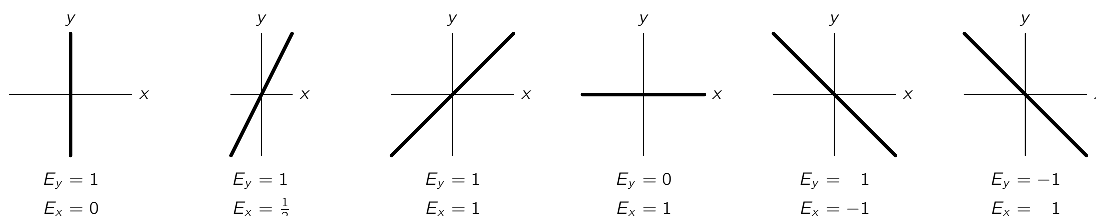


Figura 33-1. Superposición de vibraciones x y vibraciones y en fase.

Cuando la vibración x y la vibración y no están en fase, el vector campo eléctrico se mueve en una elipse, y podemos ilustrar esto de una manera familiar. Si colgamos una bola de un soporte mediante una larga cuerda de manera que pueda oscilar libremente en un plano horizontal, ejecutará oscilaciones sinusoidales. Si imaginamos coordenadas x y y horizontales con sus orígenes en la posición de reposo de la bola, ésta puede oscilar sea en la dirección x o la dirección y con la misma frecuencia pendular. Seleccionando el desplazamiento inicial y la velocidad inicial apropiados, podemos poner la bola en oscilación, sea a lo largo del eje x o del eje y, o a lo largo de cualquier línea recta en el plano xy. Estos movimientos de la bola son análogos a las oscilaciones del vector eléctrico ilustrado en la figura 33-1. En cada ejemplo, como las vibraciones x y las vibraciones y alcanzan sus máximos y mínimos en el mismo momento, las oscilaciones x y y están en fase. Pero ahora sabemos que el movimiento más general de la bola es el movimiento en una elipse, lo que corresponde a oscilaciones en que las direcciones x y y *no* están en la misma fase. La superposición de vibraciones x y y que no están en fase está ilustrada en la figura 33-2, para una variedad de ángulos entre la fase

de vibración x y la vibración y. El resultado general es que el vector eléctrico se mueve en una elipse. El movimiento en línea recta es un caso particular que corresponde a un desfase cero (o un múltiplo entero de π); el movimiento en círculo corresponde a amplitudes iguales con un desfase de 90° (o cualquier múltiplo entero impar de $\pi/2$).

En la figura 33-2 hemos representado los vectores de campo eléctrico en las direcciones x y y por números complejos, que son una representación conveniente para expresar el desfase. No confundan las componentes reales e imaginarias del vector eléctrico complejo en esta notación con las coordenadas x y y del campo. Las coordenadas x y y graficadas en las figuras 33-1 y 33-2 son campos eléctricos reales que podemos medir. Las componentes reales e imaginarias de un vector campo eléctrico complejo son solamente una conveniencia matemática y no tienen significado físico.

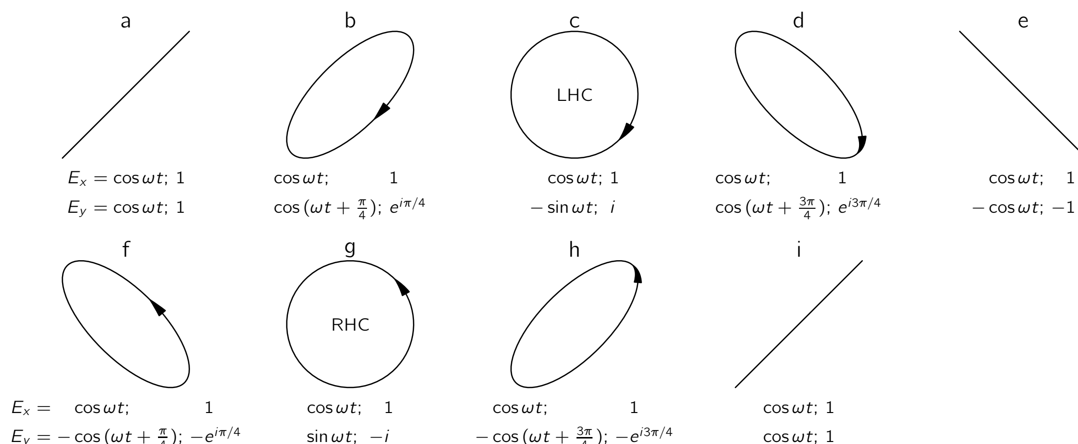


Figura 33-2. Superposición de vibraciones x y vibraciones y con amplitudes iguales, pero diferentes fases relativas. Las componentes E_x y E_y se expresan tanto en las notaciones real como compleja.

Ahora algo de terminología. La luz está *linealmente polarizada* (a veces se llama polarizada en un plano) cuando el vector eléctrico oscila en una línea recta, la figura 33-1 ilustra la polarización lineal. Cuando el extremo del vector campo eléctrico viaja en una elipse, la luz está *elípticamente polarizada*. Cuando el extremo del vector eléctrico viaja en un círculo, tenemos *polarización circular*. Si el extremo del vector eléctrico, cuando lo observamos a medida que la luz se acerca derecho hacia nosotros, se mueve en la dirección antihoraria, lo llamamos polarización circular derecha. La figura 33-2(g) ilustra la polarización circular derecha, y la figura 33-2(c) muestra la polarización circular izquierda. En ambos casos la luz sale del papel. Nuestra convención para designar la polarización circular izquierda y derecha es compatible con la que se usa hoy día para todas las otras partículas que muestran polarización en física (por ejemplo, electrones). Sin embargo, en algunos libros sobre óptica se usan convenciones opuestas, así que se debe tener cuidado.

Hemos considerado luz lineal, circular y elípticamente polarizada, lo que cubre todo excepto el caso de luz *no polarizada*. Ahora bien, ¿cómo puede la luz no estar polarizada cuando sabemos que debe vibrar en una u otra de esas elipses? Si la luz no es absolutamente monocromática, o si las fases x y y no se mantienen perfectamente juntas, de manera que el vector eléctrico vibra primero en una dirección y luego en otra, la polarización está constantemente cambiando. Recuerden que un átomo emite durante 10^{-8} s, y si un átomo emite con cierta polarización, y luego otro átomo emite luz con una polarización diferente, las polarizaciones cambiarán cada 10^{-8} s. Si la polarización cambia más rápidamente de lo que podemos detectarla, la llamamos luz no polarizada,

porque todos los efectos de la polarización se promedian a cero. Ninguno de los efectos de interferencia de la polarización se manifestaría con luz no polarizada. Pero como se desprende de la definición, la luz es no polarizada solamente si somos incapaces de averiguar si la luz está polarizada o no.

33-2 Polarización de luz dispersada

El primer ejemplo del efecto de polarización que ya hemos discutido es la dispersión de la luz. Consideremos un rayo de luz, por ejemplo, del Sol, brillando en el aire. El campo eléctrico producirá oscilaciones de cargas en el aire, y el movimiento de esas cargas irradiará luz con su máxima intensidad en un plano normal a la dirección de vibración de las cargas. El rayo proveniente del Sol no está polarizado, por lo que la dirección de polarización y la dirección de vibración de las cargas en el aire cambian constantemente. Si consideramos luz dispersada 90°, la vibración de las partículas cargadas irradia hacia el observador sólo cuando la dirección de vibración es perpendicular a la línea de visión del observador, y entonces la luz estará polarizada según la dirección de vibración. Así, la dispersión es un ejemplo de uno de los medios para producir polarización.

33-3 Birrefringencia

Otro efecto interesante de la polarización es que hay sustancias para las cuales el índice de refracción es diferente para luz linealmente polarizada en una dirección y linealmente polarizada en otra. Supongan que tengamos algún material que consista en moléculas largas, no esféricas, más largas que anchas, y supongan que esas moléculas estén dispuestas en la sustancia con sus ejes largos paralelos. Entonces, ¿qué sucede cuando el vector eléctrico oscilante pasa a través de esta sustancia? Supongamos que debido a la estructura de la molécula, los electrones de la sustancia responden más fácilmente a oscilaciones en la dirección paralela a los ejes de las moléculas que a lo que responderían si el campo eléctrico tratara de empujarlos perpendicular al eje molecular. En esta forma esperamos una respuesta diferente para polarización en una dirección que para polarización perpendicular a esa dirección. Llamemos *eje óptico* a la dirección de los ejes de la molécula. Cuando la polarización está en la dirección del eje óptico, el índice de refracción es diferente que el que sería si la dirección de polarización fuera perpendicular a él. Tal sustancia se llama *birrefringente.* Tiene dos refringencias, es decir, dos índices de refracción, que dependen de la dirección de polarización dentro de la sustancia. ¿Qué tipo de sustancia puede ser birrefringente? Una sustancia birrefringente tiene una cierta cantidad de alineamiento, por una razón u otra, de moléculas asimétricas. Naturalmente, un cristal cúbico, que tiene la simetría de un cubo, no puede ser birrefringente. Pero cristales en forma de agujas largas indudablemente contienen moléculas que son asimétricas, y se observa este efecto muy fácilmente.

Veamos qué efectos esperaríamos si tuviéramos que hacer pasar luz polarizada a través de una placa de sustancia birrefringente. Si la polarización es paralela al eje óptico, la luz atravesará con una velocidad, si la polarización es perpendicular al eje, la luz se transmite con una velocidad diferente. Una situación interesante se produce cuando, digamos, la luz está linealmente polarizada a 45° con el eje óptico. Ahora bien, la polarización a 45°, lo hemos observado ya, se puede representar como superposición de polarizaciones x y y de igual amplitud y en fase, como se muestra en la figura 33-2(a). Como las polarizaciones x y y viajan con velocidades distintas, sus fases cambian con velocidades diferentes cuando la luz atraviesa la sustancia. De esta manera, aunque al comienzo las vibraciones x y y están en fase, dentro del material, el desfase entre las vibraciones x y y es proporcional a la profundidad en la sustancia. A medida que la luz avanza a través del material, la polarización cambia en la forma que se muestra en la serie de diagramas de la figura 33-2. Si el espesor de la placa es el preciso para introducir un desfase de 90° entre las polarizaciones x y y, como en la figura 33-2(c), la luz saldrá polarizada circularmente. Ese espesor se llama placa de un cuarto de

onda, porque introduce un desfase de un cuarto de ciclo entre las polarizaciones x y y. Si se envía luz linealmente polarizada a través de dos placas de un cuarto de onda, saldrá nuevamente polarizada plana, pero perpendicularmente a la dirección original, como podemos ver en la figura 33-2(e).

Se puede ilustrar fácilmente este fenómeno con un pedazo de celofán. El celofán está hecho de moléculas largas, fibrosas, y no es isótropo puesto que las fibras yacen preferentemente en cierta dirección. Para demostrar la birrefringencia necesitamos un haz de luz linealmente polarizada, y podemos obtenerlo convenientemente haciendo pasar luz no polarizada a través de una lámina de polaroide. El polaroide, que trataremos con más detalle más tarde, tiene la propiedad útil de que transmite luz que está linealmente polarizada, paralela al eje del polaroide con muy poca absorción, pero la luz polarizada en una dirección perpendicular al eje del polaroide es absorbida fuertemente. Cuando hacemos pasar luz no polarizada a través de una lámina de polaroide, solamente pasa aquella parte del rayo no polarizado que está vibrando paralelo al eje del polaroide, por lo que el rayo transmitido está linealmente polarizado. Esta misma propiedad del polaroide es también útil para detectar la dirección de polarización de un rayo linealmente polarizado, o para determinar si un rayo está linealmente polarizado o no. Simplemente se hace pasar el rayo de luz a través de la lámina de polaroide y se rota el polaroide en el plano normal al rayo. Si el rayo está linealmente polarizado, no será transmitido a través de la lámina cuando el eje del polaroide es normal a la dirección de polarización. El rayo transmitido se atenúa sólo ligeramente cuando el eje de la lámina de polaroide se rota en 90°. Si la intensidad transmitida es independiente de la orientación del polaroide, el rayo no está linealmente polarizado.

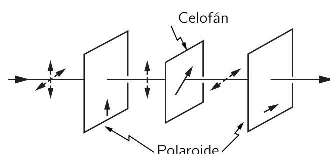


Figura 33-3. Una demostración experimental de la birrefringencia del celofán. Los vectores eléctricos de la luz se indican con líneas de trazos. Los ejes de pasada de las láminas de polaroide y los ejes ópticos del celofán se indican por flechas. El haz incidente no está polarizado.

Para demostrar la birrefringencia del celofán usamos dos láminas de polaroide, como muestra la figura 33-3. La primera nos da un rayo linealmente polarizado que hacemos pasar a través del celofán y luego a través de la segunda lámina de polaroide, la cual sirve para detectar cualquier efecto que pueda haber tenido el celofán sobre la luz polarizada que lo atraviesa. Si colocamos primero los ejes de las dos láminas de polaroide perpendiculares entre sí y quitamos el celofán, ninguna luz se transmitirá a través del segundo polaroide. Si ahora introducimos el celofán entre las dos láminas de polaroide y rotamos la lámina alrededor del eje del rayo, observamos que en general el celofán hace posible que algo de luz pase a través del segundo polaroide. Sin embargo, hay dos orientaciones de la lámina de celofán, perpendiculares entre sí, que no permiten que ninguna luz pase a través del segundo polaroide. Esas orientaciones en que la luz polarizada linealmente es transmitida a través del celofán sin efecto sobre la dirección de polarización deben ser las direcciones paralela y perpendicular al eje óptico de la lámina de celofán.

Suponemos que la luz atraviesa el celofán con dos velocidades diferentes en esas dos orientaciones distintas, pero es transmitida sin cambiar la dirección de polarización. Cuando se gira el celofán a una posición intermedia entre esas dos orientaciones, como se muestra en la figura 33-3, vemos que la luz transmitida por el segundo polaroide es brillante.

Sucede justamente que el celofán ordinario usado en paquetes comerciales es de un espesor muy cercano a media onda para la mayoría de los colores en la luz blanca. Esa lámina girará el eje de la luz linealmente polarizada en 90° si el rayo incidente linealmente polarizado forma un ángulo de 45° con el eje óptico, de modo que el rayo que emerge del celofán está vibrando entonces en la dirección correcta para pasar a través de la segunda lámina de polaroide.

Si usamos luz blanca en nuestra demostración, la lámina de celofán será del espesor apropiado de media onda solamente para una componente particular de la luz blanca, y el rayo transmitido tendrá el color de esta componente. El color transmitido depende del espesor de la lámina de celofán, y podemos variar el espesor efectivo del celofán ladeándolo de manera que la luz pase a través del celofán en forma oblicua, consecuentemente a través de una trayectoria más larga en el celofán. A medida que la lámina se ladea, el color transmitido cambia. Con celofán de diferentes espesores se pueden construir filtros que transmitirán

diferentes colores. Estos filtros tienen la interesante propiedad de que transmiten un color cuando las dos láminas de polaroide tienen sus ejes perpendiculares, y el color complementario cuando los ejes de las dos láminas polaroides son paralelos.

Otra aplicación interesante de moléculas alineadas es muy práctica. Ciertos plásticos están compuestos de moléculas muy largas y complicadas todas retorcidas entre sí. Cuando el plástico se solidifica muy cuidadosamente, las moléculas están todas retorcidas en una masa, así que hay tantas alineadas en una dirección como en otra, y el plástico no es particularmente birrefringente. Generalmente se introducen esfuerzos y deformaciones cuando el material se solidifica, por lo que el material no es perfectamente homogéneo. Sin embargo, si aplicamos tensión a una pieza de este material plástico, es como si estuviéramos estirando toda una maraña de cuerdas, y habrá más cuerdas alineadas preferencialmente paralelas a la tensión que en cualquier otra dirección. Por lo cual, cuando se aplica un esfuerzo a ciertos plásticos, se vuelven birrefringentes, y uno puede ver los efectos de la birrefringencia haciendo pasar luz polarizada a través del plástico. Si examinamos la luz transmitida a través de una lámina polaroide, se observarán dibujos de franjas oscuras y claras (en color, si se usa luz blanca). Los dibujos se mueven cuando se aplica esfuerzo a la muestra, y contando las franjas y viendo dónde está la mayoría de ellas, se puede determinar cuánto es el esfuerzo. Los ingenieros usan este fenómeno como un medio de encontrar los esfuerzos en piezas de formas raras que son difíciles de calcular.

Otro ejemplo interesante de una forma de obtener birrefringencia es mediante una sustancia líquida. Consideren un líquido compuesto de moléculas asimétricas largas que llevan una carga media positiva o negativa cerca de los extremos de la molécula, de modo que la molécula es un dipolo eléctrico. En las colisiones en el líquido las moléculas estarán ordinariamente orientadas al azar, con tantas moléculas dirigidas en una dirección como en otra. Si aplicamos un campo eléctrico, las moléculas tenderán a alinearse, y en el momento en que se alinean, el líquido se pone birrefringente. Con dos láminas de polaroide y una celda transparente que contiene ese líquido polar, podemos inventar un dispositivo con la propiedad de que la luz es transmitida solamente cuando se aplica el campo eléctrico. Así tenemos un interruptor eléctrico para la luz, que se llama *celda de Kerr*. Este efecto, que un campo eléctrico puede producir birrefringencia en ciertos líquidos, se llama efecto Kerr.

33-4 Polarizadores

Hasta ahora hemos considerado sustancias en las cuales el índice de refracción es diferente para luz polarizada en direcciones diferentes. De un valor muy práctico son aquellos cristales y otras sustancias en las cuales no solamente el índice, sino también el coeficiente de absorción es diferente para luz polarizada en direcciones diferentes. Con los mismos argumentos que sustentaron la idea de birrefringencia, es comprensible que la absorción pueda variar con la dirección en la cual las cargas son forzadas a vibrar en una sustancia anisótropa. Por ejemplo, la turmalina y el polaroidecl. El polaroide consiste en una delgada capa de pequeños cristales de herapatita (una sal de yodo y quinina), todas alineadas con sus ejes paralelos. Éste tipo de cristal absorbe luz cuando las oscilaciones están en una dirección, y no absorben mucha cuando las oscilaciones están en otras direcciones.

Supongan que enviamos luz sobre una lámina de polaroide linealmente polarizada en un ángulo θ con la dirección de pasada. ¿Cuál intensidad pasará? Esta luz incidente puede ser resuelta en una componente según la dirección de pasada, la cual es proporcional a $\cos\theta$, y una componente perpendicular a la dirección de pasada, la cual es proporcional a $\operatorname{sen}\theta$. La amplitud que sale del polaroide es solamente la parte $\cos\theta$, la componente $\operatorname{sen}\theta$ es absorbida. La amplitud que atraviesa el polaroide es más pequeña que la amplitud que entró, en un factor $\cos\theta$. La energía que atraviesa el polaroide, es decir, la intensidad de la luz, es proporcional al cuadrado de $\cos\theta$. $\cos^2\theta$ es, entonces, la intensidad transmitida

cuando la luz polarizada entra en un ángulo θ con la dirección de pasada. La intensidad absorbida, naturalmente, es $\operatorname{sen}^2\theta$.

Se presenta una interesante paradoja en la situación siguiente. Sabemos que no es posible enviar un rayo de luz a través de dos láminas polaroides con sus ejes cruzados en ángulos rectos. Pero si colocamos una tercera lámina polaroide *entre* las primeras dos, con su eje de pasada a 45° con los ejes cruzados, algo de luz se transmite. Sabemos que el polaroide absorbe luz, no crea cosa alguna. Sin embargo, la colocación de un tercer polaroide a 45° permite que pase más luz. El análisis de este fenómeno se deja como ejercicio para el estudiante.

Uno de los ejemplos más interesantes de polarización no está en cristales complicados o sustancias difíciles, sino en una de las más simples y familiares de las situaciones —la reflexión de la luz en una superficie—. Créanlo o no, cuando la luz se refleja en una superficie de vidrio puede estar polarizada, y la explicación física de esto es muy simple. Fue descubierto empíricamente por Brewster que la luz reflejada en una superficie está completamente polarizada si el rayo reflejado y el rayo refracción dentro del material forman un ángulo recto. La situación se ilustra en la figura 33-4. Si el rayo incidente está polarizado en el plano de incidencia, no habrá reflexión en absoluto. Sólo si el rayo incidente está polarizado normalmente al plano de incidencia será reflejado. Es muy fácil entender la razón. En el material reflector la luz está polarizada transversalmente, y sabemos que es el movimiento de las cargas en el material lo que genera el rayo emergente, al cual llamamos rayo reflejado. La causa de la así llamada luz reflejada no es simplemente que el rayo incidente sea reflejado; nuestra comprensión más profunda de este fenómeno nos dice que el rayo incidente provoca una oscilación de las cargas en el material, lo que a su vez genera el rayo reflejado. En la 33-4 es claro que solamente las oscilaciones normales al papel pueden radiar en la dirección de reflexión, y consecuentemente el rayo reflejado estará polarizado normalmente al plano de incidencia. Si el rayo incidente está polarizado en el plano de incidencia, no habrá luz reflejada.

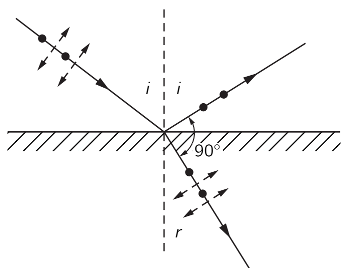


Figura 33-4. Reflexión de luz linealmente polarizada en el ángulo de Breswter. La dirección de polarización se indica por línea discontinua; los circulitos indican polarización normal al papel.

Este fenómeno se demuestra fácilmente reflejando un rayo de luz linealmente polarizada en un pedazo de vidrio plano. Si el vidrio se gira para presentar diferentes ángulos de incidencia al rayo polarizado, se ve atenuación aguda de la intensidad reflejada cuando el ángulo de incidencia pasa a través del ángulo de Brewster. Esta atenuación se observa solamente si el plano de polarización yace en el plano de incidencia. Si el plano de polarización es normal al plano de incidencia, la intensidad reflejada usual se observa en todos los ángulos.

33-5 Actividad óptica

Otro notabilísimo efecto de la polarización se observa en materiales compuestos de moléculas que no tienen simetría de reflexión: moléculas semejantes a un sacacorchos, o como una mano enguantada, o de cualquier forma que mirada a través de un espejo estaría invertida en la misma forma que un guante izquierdo se refleja como un guante derecho. Supongamos que todas las moléculas de la sustancia son iguales, dicho de otro modo, ninguna es una imagen especular de otra. Esa sustancia puede mostrar un interesante efecto llamado actividad óptica, en el cual, cuando la luz polarizada pasa a través de la sustancia, la dirección de polarización rota alrededor del eje del rayo.

Para comprender el fenómeno de actividad óptica se requiere algún cálculo, pero podemos ver cualitativamente cómo se produce el efecto sin hacer realmente cálculos. Considere una molécula asimétrica en la forma de una espiral, véase la figura 33-5. Las moléculas no necesitan realmente tener formas como un sacacorchos para exhibir actividad óptica, pero ésta es una forma simple que tomaremos como ejemplo típico de las que no tienen simetría de reflexión. Cuando un rayo de luz linealmente polarizado según la dirección y cae sobre esta molécula, el campo eléctrico forzará las cargas hacia arriba y hacia abajo de la hélice, generando así una corriente en la dirección y e irradiando un campo eléctrico E_y polarizado en la dirección y. Sin embargo, si los electrones están constreñidos a moverse a lo largo de la espiral, también deben moverse en la dirección x cuando

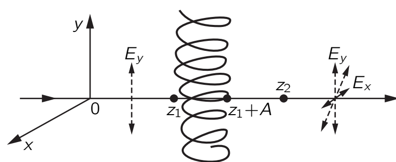


Figura 33-5. Una molécula de forma tal que no es simétrica bajo reflexión en un espejo. Un rayo de luz, linealmente polarizado en la dirección y, incide sobre la molécula.

son forzados hacia arriba y hacia abajo. Cuando una corriente fluye hacia encima de la espiral, también está fluyendo hacia dentro del papel en $z = z_1$, y hacia fuera del papel en $z = z_1 + A$, si A es el diámetro de nuestra espiral molecular. Se puede suponer que la corriente en la dirección x no producirá radiación neta puesto que las corrientes están en direcciones opuestas en lados opuestos de la espiral. Sin embargo, si consideramos las componentes x del campo eléctrico que llegan a $z = z_2$, vemos que el campo irradiado por la corriente en $z = z_1 + A$ y el campo irradiado en $z = z_1$ llegan a z_2 separados en el tiempo por la cantidad A/c, y por lo tanto separados en fase en $\pi + \omega A/c$. Como la diferencia de fase no es exactamente π, los dos campos no se anulan exactamente, y nos queda una pequeña componente x en el campo eléctrico generado por el movimiento de los electrones en la molécula, mientras que el campo eléctrico excitador tenía solamente una componente y. Esta pequeña componente x, sumada a la gran componente y, produce un campo resultante que está inclinado ligeramente respecto al eje y, la dirección original de polarización. A medida que la luz se mueve a través del material, la dirección de polarización rota alrededor del eje del rayo. Dibujando unos pocos ejemplos y considerando las corrientes que un campo eléctrico incidente pondrá en movimiento, uno se puede convencer de que la existencia de la actividad óptica y el signo de rotación son independientes de la orientación de las moléculas.

La miel de maíz es una sustancia común que posee actividad óptica. El fenómeno se demuestra fácilmente con una lámina de polaroide para producir un rayo linealmente polarizado, una celda de transmisión que contiene miel de maíz, y un segundo polaroide para detectar la rotación de la dirección de polarización a medida que la luz pasa a través del jarabe de maíz.

33-6 Intensidad de la luz reflejada

Consideremos ahora cuantitativamente el coeficiente de reflexión en función del ángulo. La figura 33-6(a) muestra un rayo de luz incidiendo sobre una superficie de vidrio, donde se refleja y se refracta parcialmente en el vidrio. Supongamos que el rayo incidente, de amplitud unitaria, está linealmente polarizado normal al plano del papel. Llamaremos b a la amplitud de la onda reflejada, y a a la amplitud de la onda refractada. Las ondas refractadas y reflejadas estarán, por supuesto, linealmente polarizadas, y los vectores campo eléctrico de las ondas incidente, reflejada y refractada son todos paralelos entre sí. La figura 33-6(b) muestra la misma situación, pero ahora suponemos que la onda incidente, de amplitud unitaria, está polarizada en el plano de papel. Llamemos ahora B y A a las amplitudes de las ondas reflejada y refractada, respectivamente.

Deseamos calcular la intensidad de la reflexión en las dos situaciones ilustradas en las figuras 33-6(a) y (b). Sabemos que cuando el ángulo entre el rayo reflejado y el rayo refractado es un ángulo recto, no habrá onda reflejada en la figura 33-6(b), pero veamos si no podemos obtener una respuesta cuantitativa, una fórmula exacta para B y b en función del ángulo de incidencia i.

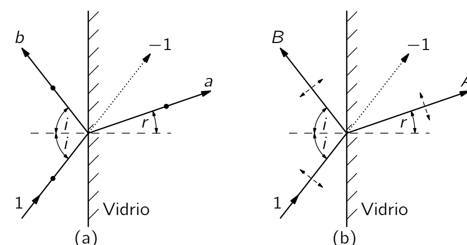


Figura 33-6. Una onda incidente de amplitud unitaria se refleja y refracta en una superficie de vidrio. En (a) la onda incidente está linealmente polarizada normalmente al plano del papel. En (b) la onda incidente está linealmente polarizada en la dirección indicada por la línea discontinua.

El principio que debemos comprender es el siguiente. Las corrientes que se generan en el vidrio producen dos ondas. Primero, producen la onda reflejada. Además, sabemos que si no hubiera corrientes generadas en el vidrio, la onda incidente continuaría directamente hacia el interior del vidrio. Recuerden que todas las fuentes en el mundo forman el campo resultante. La fuente del rayo de luz incidente produce un campo de amplitud unitaria que se movería hacia el interior del vidrio a lo largo de la línea punteada en la figura. Este campo no se observa y, por lo tanto, las corrientes generadas en el vidrio deben producir un campo de amplitud -1 que se mueve a lo largo de la línea punteada. Usando esto, calcularemos la amplitud de las ondas refractadas a y A.

En la figura 33-6(a) vemos que el campo de amplitud b es irradiado por el movimiento de cargas dentro del vidrio que están respondiendo a un campo a dentro del vidrio, y que por lo tanto b es proporcional a a. Podríamos suponer, puesto que nuestras dos figuras son exactamente iguales, excepto para la dirección de polarización, que el cociente B/A sería la misma que la razón b/a. Esto no es

completamente cierto, sin embargo, porque en la figura 33-6(b) las direcciones de polarización no son todas paralelas entre sí, como lo son en la figura 33-6(a). Es sólo la componente del campo eléctrico en el vidrio la que es perpendicular a B, $A\cos(i+r)$, la que es efectiva en producir B. La expresión correcta para la proporcionalidad es:

$$\frac{b}{a} = \frac{B}{A\cos(i+r)}. \tag{33.1}$$

Ahora usamos un truco. Sabemos que tanto en (a) como en (b) de la figura 33-6, el campo eléctrico en el vidrio debe producir oscilaciones de las cargas, que generan un campo eléctrico de amplitud -1, polarizado paralelo al rayo incidente, y moviéndose en la dirección de la línea punteada. Pero vemos en la parte (b) de la figura que solamente la componente del campo eléctrico en el vidrio, que es normal a la línea discontinua, tiene la polarización correcta para producir este campo, mientras que en la figura 33-6(a) la amplitud completa a es efectiva, puesto que la polarización de la onda a es paralela a la polarización de la onda de amplitud -1. Por lo tanto, podemos escribir

$$\frac{A\cos(i-r)}{a} = \frac{-1}{-1}, \tag{33.2}$$

puesto que las amplitudes en el primer miembro de la ecuación (33.2) producen cada una la onda de amplitud -1.

Dividiendo la ecuación (33.1) entre (33.2), obtenemos:

$$\frac{B}{b} = \frac{\cos(i+r)}{\cos(i-r)}, \tag{33.3}$$

un resultado que podemos comparar con el que ya conocemos. Si hacemos $(i+r) = 90°$, la ecuación (33.3) da $B = 0$, como Brewster dice que debería ser, por lo que nuestros resultados hasta ahora por lo menos no están evidentemente equivocados.

Hemos supuesto amplitudes unitarias para las ondas incidentes, así que $|B|^2/1^2$ es el coeficiente de reflexión para ondas polarizadas en el plano de incidencia, y $|b|^2/1^2$ es el coeficiente de reflexión para ondas polarizadas perpendicularmente al plano de incidencia. El cociente de estos dos coeficientes de reflexión está determinado por la ecuación (33.3).

¡Ahora realizamos un milagro, y calculamos no sólo el cociente, sino cada coeficiente $|B|^2$ y $|b|^2$ individualmente! Por la conservación de la energía, se tiene que la energía de la onda refractada debe ser igual a la energía incidente menos la energía de la onda reflejada, $1-|B|^2$ en un caso, $1-|b|^2$ en el otro. Además, la energía que pasa al interior del vidrio en la figura 33-6(b) es a la energía que pasa al vidrio en la figura 33-6(a), como la razón de los cuadrados de las amplitudes refractadas, $|A|^2/|a|^2$. Uno podría preguntar si realmente sabemos cómo calcular la energía en el interior del vidrio, porque, después de todo, hay energías de movimientos de los átomos además de la energía del campo eléctrico. Pero es evidente que la totalidad de las diversas contribuciones a la energía total será proporcional al cuadrado de la amplitud del campo eléctrico. Por lo tanto, podemos escribir,

$$\frac{1-|B|^2}{1-|b|^2} = \frac{|A|^2}{|a|^2}. \tag{33.4}$$

Ahora sustituimos la ecuación (33.2) para eliminar A/a de la expresión anterior, y expresamos B en función de b mediante la ecuación (33.3)

$$\frac{1-|b|^2\dfrac{\cos^2(i+r)}{\cos^2(i-r)}}{1-|b|^2} = \frac{1}{\cos^2(i-r)}. \tag{33.5}$$

Esta ecuación contiene solamente una amplitud desconocida, b. Despejando $|b|^2$ obtenemos:

$$|b|^2 = \frac{\operatorname{sen}^2(i-r)}{\operatorname{sen}^2(i+r)} \tag{33.6}$$

y con la ayuda de (33.3),

$$|B|^2 = \frac{\tan^2(i-r)}{\tan^2(i+r)}. \tag{33.7}$$

¡Así, hemos encontrado el coeficiente de reflexión $|b|^2$ para una onda incidente polarizada perpendicularmente al plano de incidencia, y también el coeficiente de reflexión $|B|^2$ para una onda incidente polarizada en el plano de incidencia!

Es posible seguir con argumentos de esta naturaleza y deducir que b es real. Para demostrar esto, debe considerarse un caso donde la luz llega desde ambos lados de la superficie del vidrio en el mismo instante, una situación no fácil de disponer experimentalmente, pero divertida de analizar teóricamente. Si analizamos este caso general, podemos probar que b debe ser real, y por lo tanto, de hecho, que $b = \pm \operatorname{sen}(i-r)/\operatorname{sen}(i+r)$. Es aún posible determinar el signo considerando el caso de una lámina, muy, muy delgada, en que hay reflexión desde las superficies frontal y trasera, y calculando cuánta luz se refleja. Sabemos cuánta luz debería ser reflejada por una capa delgada, porque sabemos cuánta corriente se genera, y hasta hemos determinado los campos producidos por éstas.

Se puede demostrar por estos argumentos que:

$$b = -\frac{\operatorname{sen}(i-r)}{\operatorname{sen}(i+r)}, \qquad B = -\frac{\tan(i-r)}{\tan(i+r)}. \tag{33.8}$$

Estas expresiones para los coeficientes de reflexión en función de los ángulos de incidencia y de refracción se llaman fórmulas de reflexión de Fresnel.

Si consideramos el límite cuando los ángulos i y r se aproximan a cero, encontramos, para el caso de incidencia normal, que $B^2 \approx b^2 \approx (i-r)^2/(i+r)^2$, para ambas polarizaciones, puesto que los senos prácticamente son iguales a los ángulos, como también lo son las tangentes. Pero sabemos que $\operatorname{sen} i/\operatorname{sen} r = n$, y cuando los ángulos son pequeños, $i/r \approx n$. Así, pues, es fácil demostrar que el coeficiente de reflexión para incidencia normal es:

$$B^2 = b^2 = \frac{(n-1)^2}{(n+1)^2}.$$

Es interesante determinar cuánta luz se refleja en incidencia normal en la superficie del agua, por ejemplo. Para el agua, n es 4/3, así que el coeficiente de reflexión es $(1/7)^2 \approx 2\,\%$. En incidencia normal, solamente un 2 % de la luz se refleja en una superficie de agua.

33-7 Refracción anómala

El último efecto de polarización que consideraremos fue efectivamente uno de los primeros en ser descubierto: la refracción anómala. Unos marineros que visitaron Islandia trajeron de regreso a Europa cristales de espato de Islandia ($CaCO_3$) que tenían la divertida propiedad de hacer que cualquier cosa vista a través del cristal apareciera duplicada, es decir, como dos imágenes. Esto llamó la atención de Huygens, y desempeño un papel importante en el descubrimiento de la polarización. Como usualmente sucede, los fenómenos que se descubren primero son los más difíciles, en última instancia, de explicar. Es solamente después que comprendemos enteramente el concepto físico, que podemos seleccionar cuidadosamente aquellos fenómenos que demuestran el concepto más claro y simple.

La refracción anómala es un caso particular de la misma birrefringencia que consideramos anteriormente. La refracción anómala se produce cuando el eje óptico, el eje largo de nuestras moléculas asimétricas, *no* es paralelo a la superficie del cristal. En la figura 33-7 están dibujadas dos piezas de material birrefringente,

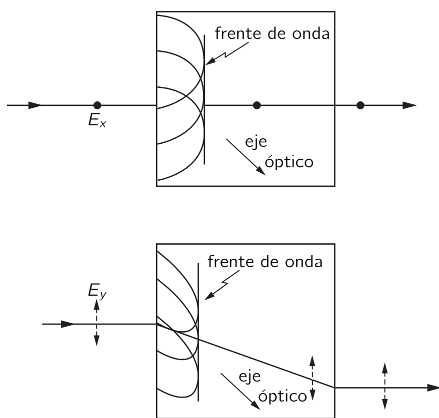


Figura 33-7. El diagrama superior muestra la trayectoria del rayo ordinario a través de un cristal birrefringente. El rayo extraordinario se muestra en el diagrama inferior. El eje óptico yace en el plano del papel.

con el eje óptico, como se muestra. En la figura superior, el rayo incidente que cae sobre el material está linealmente polarizado en una dirección perpendicular al eje óptico del material. Cuando este rayo choca con la superficie del material, cada punto de la superficie actúa como una fuente de una onda que viaja dentro del cristal con velocidad $v_\perp$, la velocidad de la luz en el cristal cuando el plano de polarización es normal al eje óptico. El frente frente de es justamente la envolvente o lugar geométrico de todas estas pequeñas ondas esféricas, y este frente de onda se mueve derecho a través del cristal, y sale por el otro lado. Éste es precisamente el comportamiento ordinario que esperaríamos, y este rayo se llama *rayo ordinario*.

En la figura inferior, la luz linealmente polarizada que incide sobre el cristal tiene su dirección de polarización girada en 90°, así que el eje óptico yace en el plano de polarización. Si ahora consideramos las pequeñas ondas que se originan en cualquier punto de la superficie del cristal, vemos que no se esparcen como ondas esféricas. La luz viajando a lo largo del eje óptico viaja con velocidad $v_\perp$, porque la polarización es perpendicular al eje óptico, mientras que la luz viajando perpendicular al eje óptico viaja con velocidad $v_\parallel$, porque la polarización es paralela al eje óptico. En un material birrefringente $v_\parallel \neq v_\perp$, y en la figura $v_\parallel < v_\perp$. Un análisis más completo demostrará que las ondas se esparcen en la superficie de un elipsoide, con el eje óptico como eje mayor del elipsoide. La envolvente de todas esas pequeñas ondas elípticas es el frente de onda que avanza a través del cristal en la dirección mostrada. De nuevo, en la superficie posterior, el rayo será reflectado justamente como lo fue en la superficie frontal, así que la luz emerge paralela al rayo incidente, pero desplazada de éste. Claramente, este rayo no sigue la ley de Snell, sino que sigue en una dirección extraordinaria. Por lo tanto, se le llama *rayo extraordinario*.

Cuando un rayo no polarizado incide sobre un cristal refractante anómalo, se separa en un rayo ordinario, que atraviesa directamente de la manera normal, y en un rayo extraordinario que se desplaza a medida que atraviesa el cristal. Esos dos rayos emergentes están linealmente polarizados perpendicularmente entre sí. Que esto sea verdad se puede demostrar fácilmente con una lámina de polaroide para analizar la polarización de los rayos emergentes. Podemos también demostrar que nuestra interpretación de este fenómeno es correcta, enviando luz linealmente polarizada dentro del cristal. Mediante una orientación adecuada de la dirección de polarización del rayo incidente, podemos hacer que esta luz atraviese directamente sin desdoblamiento, o podemos hacer que atraviese sin desdoblamiento pero con un desplazamiento.

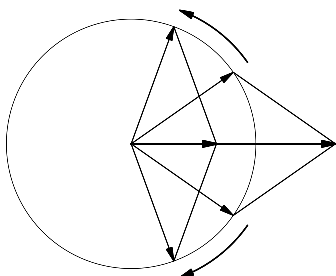


Figura 33-8. Dos vectores de igual amplitud rotando en sentido opuesto se suman para producir un vector en una dirección fija, pero con amplitud oscilante.

Hemos representado todos los casos diferentes de polarización en las figuras 33-1 y 33-2, como superposiciones de dos casos especiales de polarización, específicamente x y y, en cantidades y fases diversas. Otros pares podrían haberse usado igualmente bien. Polarización según cualquier par de ejes x' y y' perpendiculares, inclinados respecto a x y y, servirían igualmente; por ejemplo, cualquier polarización puede estar formada de la superposición de casos (a) y (e) de la figura 33-2. Es interesante, sin embargo, que esta idea puede ser ampliada también a otros casos. Por ejemplo, cualquier polarización *lineal* puede estar formada por la superposición de cantidades adecuadas con fases adecuadas de polarizaciones *circulares* derecha e izquierda; casos (c) y (g) de la figura 33-2, puesto que dos vectores iguales rotando en direcciones opuestas se suman para dar un vector único oscilando en una línea recta (figura 33-8). Si la fase de uno está desplazada respecto del otro, la línea es inclinada. Así, pues, todos los dibujos de la figura 33-1 podrían rotularse como "la superposición de cantidades iguales de luz circularmente polarizada derecha e izquierda con diferentes fases relativas". Cuando la componente izquierda se corre detrás de la derecha en fase, la dirección de la polarización lineal cambia. Por lo tanto, los materiales ópticamente activos son, en un sentido, birrefringentes. Sus propiedades pueden ser descritas diciendo que tienen diferentes índices para luz circularmente polarizada derecha e izquierda. La superposición de luz circularmente polarizada derecha e izquierda de diferentes intensidades produce luz elípticamente polarizada.

La luz circularmente polarizada tiene otra propiedad interesante —transporta *momentum angular*, respecto a la dirección de propagación—. Para ilustrar esto, supongamos que esa luz cae sobre un átomo representado por un oscilador armónico que se puede desplazar igualmente bien en cualquier dirección en el plano xy. Entonces, el desplazamiento x del electrón responderá a la componente E_x del campo mientras que la componente y responde, igualmente, a la componente E_y del campo pero retrasado en una fase de 90°. Esto es, el electrón que responde gira en un círculo con velocidad angular ω, en respuesta al campo eléctrico rotatorio de la luz (figura 33-9). Dependiendo de las características de amortiguamiento de la respuesta del oscilador, la dirección del desplazamiento $\boldsymbol{a}$ de electrón y la dirección de la fuerza $q_e\boldsymbol{E}$ sobre él no necesitan ser las mismas, sino que rotan juntas. El $\boldsymbol{E}$ puede tener una componente perpendicular $\boldsymbol{a}$, de modo que se hace trabajo sobre el sistema y se ejerce un torque r. El trabajo hecho por segundo es $\tau\omega$. En un periodo de tiempo T, la energía absorbida es $\tau\omega T$, mientras que τT es el momentum angular cedido a la materia que absorbe la energía. Vemos, por lo tanto, que *un rayo de luz circularmente polarizado que contiene una energía total* $\mathcal{E}$ *transporta un momentum angular (con vector dirigido según la dirección de propagación)* $\mathcal{E}/\omega$. Porque cuando este rayo es absorbido se cede ese momentum al absorbente. La luz circularmente polarizada a izquierda transporta momentum angular de signo opuesto, $-\mathcal{E}/\omega$.

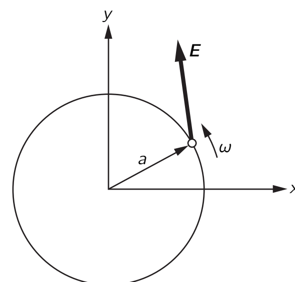


Figura 33-9. Una carga moviéndose en un círculo en respuesta a luz circularmente polarizada.

34

Efectos relativistas en la radiación

34-1 Fuentes en movimiento

En este capítulo describiremos varios efectos misceláneos relacionados con la radiación, y entonces habremos terminado con la teoría clásica de la propagación de la luz. En nuestro análisis de la luz hemos ido bastante lejos y con considerable detalle. El único fenómeno de alguna consecuencia asociado con la radiación electromagnética que no hemos discutido es: qué sucede si ondas de radio están contenidas en una caja con paredes reflectoras, siendo el tamaño de la caja comparable con una longitud de onda, o si éstas son transmitidas en un tubo largo. Los fenómenos de los así llamados *resonadores de cavidad* y *guías de ondas* serán discutidos más tarde; usaremos primero otro ejemplo físico —el sonido— y luego volveremos a este tema. Este capítulo, excepto en esto, es nuestra última consideración de la teoría clásica de la luz.

Podemos resumir todos los efectos que discutiremos ahora, haciendo notar que ellos tienen que ver con los efectos de *fuentes de movimiento.* Ya no supondremos más que la fuente está localizada, con todo su movimiento a una velocidad relativamente baja alrededor de un punto fijo.

Recordemos que las leyes fundamentales de la electrodinámica dicen que, a grandes distancias de una carga móvil, el campo eléctrico está dado por la fórmula:

$$\boldsymbol{E} = -\frac{q}{4\pi\epsilon_0 c^2}\,\frac{d^2\boldsymbol{e}_{R'}}{dt^2}. \tag{34.1}$$

La segunda derivada del vector unitario $\boldsymbol{e}_{R'}$, que apunta en la dirección aparente de la carga, es el aspecto determinante del campo eléctrico. Además, este vector unitario no apunta hacia la posición *presente* de la carga, naturalmente, sino más bien en la dirección en que la carga parecería estar, si la información viaja solamente con la velocidad finita c desde la carga al observador.

Asociado con el campo eléctrico hay un campo magnético, siempre perpendicular al campo eléctrico y perpendicular a la dirección aparente de la fuente, dado por la fórmula.

$$\boldsymbol{B} = -\boldsymbol{e}_{R'} \times \boldsymbol{E}/c. \tag{34.2}$$

Hasta ahora hemos considerado solamente el caso en que los movimientos son a velocidades no relativistas; así que no hay movimiento apreciable a considerarse en la dirección de la fuente. Ahora seremos más generales y estudiaremos el caso en que el movimiento tiene una velocidad arbitraria, y veremos qué efectos diferentes pueden esperarse en aquellas circunstancias. Dejaremos que el movimiento tenga una velocidad arbitraria, pero naturalmente supondremos aún que el detector está muy lejos de la fuente.

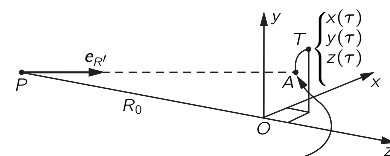


Figura 34-1. La trayectoria de una carga móvil. La verdadera posición en el tiempo τ está en T, pero la posición retardada está en A.

Ya sabemos, por nuestra discusión en el capítulo 28, que las únicas cosas que cuentan en $d^2\boldsymbol{e}_{R'}/dt^2$ son los cambios en la *dirección* de $\boldsymbol{e}_{R'}$. Sean (x, y, z) las coordenadas de la carga, con z medida a lo largo de la dirección de observación (figura 34-1). En un momento dado en el tiempo, digamos el instante τ, las tres componentes de la posición son $x(\tau)$, $y(\tau)$ y $z(\tau)$. La distancia R es muy aproximadamente igual a $R(\tau) = R_0 + z(\tau)$. Ahora bien, la dirección del vector unitario $\boldsymbol{e}_{R'}$ depende principalmente de x y y, pero muy poco de z; las

componentes transversales del vector unitario son x/R y y/R, y cuando derivamos esas componentes, obtenemos términos como R^2 en el denominador:

$$\frac{d(x/R)}{dt} = \frac{dx/dt}{R} - \frac{dz}{dt}\,\frac{x}{R^2}.$$

Así, cuando estamos suficientemente lejos, los únicos términos de los que tenemos que preocuparnos son las variaciones de x y y. Así, pues, eliminamos el factor R_0 y obtenemos

$$E_x = -\frac{q}{4\pi\epsilon_0 c^2 R_0}\,\frac{d^2x'}{dt^2},$$
$$E_y = -\frac{q}{4\pi\epsilon_0 c^2 R_0}\,\frac{d^2y'}{dt^2}, \tag{34.3}$$

donde R_0 es más o menos la distancia a q; tomémosla como la distancia OP al origen de las coordenadas (x, y, z). Entonces, el campo eléctrico es una constante multiplicada por una cosa muy simple: las segundas derivadas de las coordenadas x y y. (Podríamos expresarlo más matemáticamente, llamando x y y las componentes *transversales* del vector de posición $\boldsymbol{r}$ de la carga, pero esto no agrega mayor claridad.)

Por supuesto, nos damos cuenta de que las coordenadas deben ser medidas en el tiempo retardado. Aquí encontramos que $z(\tau)$ *sí* afecta al retardo. ¿Cuál es el tiempo retardado? Si se llama t al tiempo de observación (el tiempo en P), entonces el tiempo τ al cual esto corresponde en A, no es el tiempo t, pero se retrasa por la distancia total que la luz tiene que viajar, dividido entre la velocidad de la luz. En primera aproximación, este retardo es R_0/c, una constante (una característica no interesante), pero en la aproximación siguiente debemos incluir los efectos de la posición en la dirección z en el tiempo τ, porque si q está un poco alejada hacia atrás, hay un retardo un poco mayor. Éste es un efecto que hemos despreciado anteriormente, y es el único cambio que se necesita para hacer válidos nuestros resultados para todas las velocidades.

Lo que debemos hacer ahora es escoger cierto valor t y calcular el valor τ a partir de él, y encontrar dónde estaban x y y en aquel τ. Éstos son entonces los x y y retardados, que llamaremos x' y y', cuyas derivadas segundas determinan el campo. Así pues, τ está determinado por

$$t = \tau + \frac{R_0}{c} + \frac{z(\tau)}{c}$$

y

$$x'(t) = x(\tau), \qquad y'(t) = y(\tau). \tag{34.4}$$

Éstas son ecuaciones complicadas, pero es suficientemente fácil hacer un cuadro geométrico para describir su solución. Este cuadro nos dará una buena idea cualitativa de cómo funcionan las cosas, pero aún se requiere mucha matemática detallada para deducir los resultados precisos de un problema complicado.

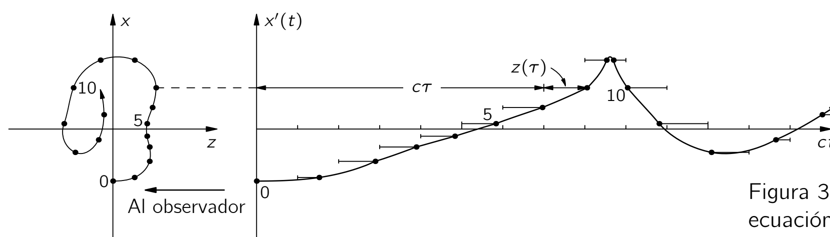


Figura 34-2. Una solución geométrica de la ecuación (34.5) para encontrar $x'(t)$.

34-2 Un modo de hallar el movimiento "aparente"

La ecuación anterior tiene una simplificación interesante. Si despreciamos el retardo constante R_0/c, que no interesa, que sólo significa que debemos cambiar el origen de t en una constante, entonces esto dice que:

$$ct = c\tau + z(\tau), \qquad x' = x(\tau), \qquad y' = y(\tau). \tag{34.5}$$

Ahora necesitamos encontrar x' y y' como funciones de t, no de τ, y podemos hacer esto de la manera siguiente: la ecuación (34.5) dice que deberíamos tomar el movimiento real y agregar una constante (la velocidad de la luz) multiplicada por τ. Lo que esto quiere decir está mostrado en la figura 34-2. Tomamos el movimiento real de la carga (mostrado a la izquierda) y nos imaginamos que a medida que se mueve está siendo arrastrada desde el punto P con la velocidad de la luz c (no hay contracciones relativistas o cualquier cosa como ésta; esto es sólo una suma matemática de los $c\tau$). En esta forma obtenemos un nuevo movimiento donde la coordenada según la visual es ct como se muestra a la derecha. (La figura muestra el resultado para un movimiento bastante complicado en un plano, pero por supuesto el movimiento puede no estar en un plano; puede ser aún más complicado que el movimiento en un plano.) La cuestión es que ahora la distancia horizontal (es decir, según la visual) no es más la antigua z, sino $z + c\tau$, y por lo tanto es ct. ¡Así que hemos encontrado una representación de la curva, x' (y y') en función de t! Todo lo que tenemos que hacer para encontrar el campo es buscar la aceleración de esta curva, o sea, derivar dos veces. Así, pues, la respuesta final es: para encontrar el campo eléctrico de una carga móvil tómese el movimiento de la carga y trasládelo hacia atrás con la velocidad c para "abrirlo", luego la curva así dibujada es una curva de las posiciones x' y y' en función de t. La aceleración de esta curva da el campo eléctrico en función de t. O bien, si queremos, podemos ahora imaginar que toda esta curva "rígida" se mueve hacia adelante con velocidad c a través del plano visual, de manera que el punto de intersección con el plano visual tiene las coordenadas x' y y'. La aceleración de este punto produce el campo eléctrico. Esta solución es tan exacta como la fórmula de que partimos —es simplemente una representación geométrica—.

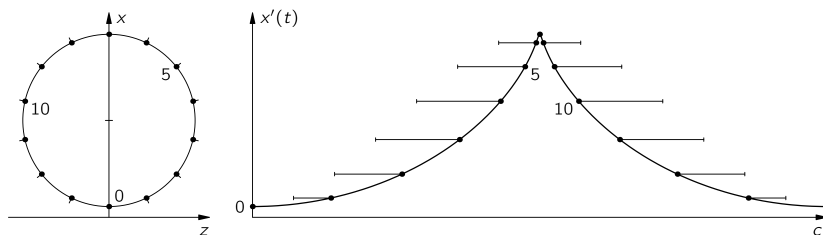


Figura 34-3. La curva $x'(t)$ para una partícula que se mueve con velocidad constante $v = 0.94\,c$ en un círculo.

Si el movimiento es relativamente lento, si tenemos, por ejemplo, un oscilador que sólo sube y baja lentamente, entonces cuando disparamos ese movimiento con la velocidad de la luz obtendríamos por supuesto una simple curva coseno, y eso da una fórmula que habíamos estado buscando por mucho tiempo: da el campo producido por una carga oscilante. Un ejemplo más interesante es un electrón moviéndose rápidamente, casi a la velocidad de la luz, en un círculo. Si miramos en el plano del círculo, el $x'(t)$ retardado aparece como muestra la figura 34-3. ¿Qué es esta curva? Si imaginamos un vector radio desde el centro del círculo a la carga y extendemos esta línea radial un poco más allá de la carga sólo un poco si se mueve rápido, entonces llegamos a un punto de la línea que va a la velocidad de la luz. Por lo tanto, cuando trasladamos este movimiento hacia atrás con la velocidad de la luz, eso corresponde a tener una rueda con una carga sobre ella rodando hacia atrás (sin deslizar) con la velocidad de la luz; así encontramos una curva que es muy parecida a una cicloide —se llama *cicloide acortada*—. Si la carga va con una velocidad muy cercana a la de la luz, las "cúspides" son en verdad muy agudas; si fuera a exactamente la velocidad de la luz, serían cúspides reales infinitamente agudas. "Infinitamente agudas" es interesante, significa que cerca de una cúspide, la segunda derivada es enorme. Una vez en cada ciclo obtenemos un pulso agudo de campo eléctrico. Esto no es todo lo que obtendríamos de un movimiento no relativista, donde cada vez

que la carga da vueltas hay una oscilación que es aproximadamente la misma "fuerza" todo el tiempo. En cambio hay pulsos muy agudos de campo eléctrico a intervalos de tiempo T_0, donde T_0 es el periodo de revolución. Estos intensos campos eléctricos son emitidos en un estrecho cono en la dirección del movimiento de la carga. Cuando la carga se aleja de P, hay muy poca curvatura y hay muy poco campo irradiado en la dirección de P.

34-3 Radiación sincrotrónica

En el sincrotrón tenemos electrones muy rápidos moviéndose en trayectorias circulares, están viajando casi a la velocidad c ¡y es posible ver la radiación antedicha como *luz* real! Discutamos esto con mayor detalle.

En el sincrotrón tenemos electrones que dan vueltas en círculos en un campo magnético uniforme. Veamos primero por qué ellos van en círculos. Conforme a la ecuación (28.2), sabemos que la fuerza sobre una partícula en un campo magnético está dada por:

$$\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}, \tag{34.6}$$

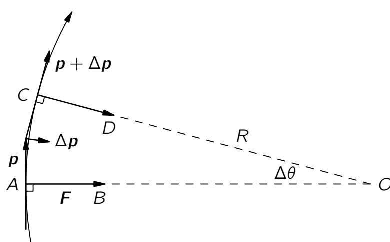


Figura 34-4. Una partícula cargada se mueve en una trayectoria circular (o helicoidal) en un campo magnético uniforme.

y es perpendicular al campo y a la velocidad. Como siempre, la fuerza es igual a la derivada del momentum respecto al tiempo. Si el campo está dirigido hacia arriba saliendo del papel, el momentum de la partícula y la fuerza sobre ésta son como muestra la figura 34-4. Como la fuerza es perpendicular a la velocidad, la energía cinética, y por lo tanto la velocidad, permanece *constante*. Todo lo que hace el campo magnético es cambiar la *dirección del movimiento*. En un corto tiempo Δt, el vector momentum cambia en ángulo recto a sí mismo en una cantidad $\Delta\boldsymbol{p} = \boldsymbol{F}\Delta t$, y por lo tanto $\boldsymbol{p}$ gira en un ángulo $\Delta\theta = \Delta p/p = qvB\Delta t/p$, puesto que $|\boldsymbol{F}| = qvB$. Pero en este mismo tiempo la partícula ha viajado una distancia $\Delta s = v\Delta t$. Evidentemente, las dos líneas AB y CD se intersecarán en un punto O tal que $OA = OC = R$, donde $\Delta s = R\Delta\theta$. Combinando esto con las expresiones anteriores, encontramos $R\Delta\theta/\Delta t = R\omega = v = qvBR/p$, de donde encontramos:

$$p = qBR \tag{34.7}$$

y

$$\omega = qvB/p. \tag{34.8}$$

Como este mismo razonamiento se puede aplicar durante el instante siguiente, el que sigue, y así sucesivamente, concluimos que la partícula debe estar moviéndose en un *círculo* de radio R con velocidad angular ω.

El resultado de que el momentum de la partícula sea igual a la carga por el radio por el campo magnético, es una ley muy importante que se usa mucho. Es importante para fines prácticos, porque si tenemos partículas elementales que tienen todas la misma carga y las observamos en un campo magnético, podemos medir los radios de curvatura de sus órbitas y, conociendo el campo magnético, determinar así el momentum de las partículas. Si multiplicamos ambos lados de la ecuación (34.7) por c, y expresamos q en términos de la carga electrónica, podemos medir el momentum en unidades *electronvoltio*. En esas unidades nuestra fórmula es:

$$pc(\text{eV}) = 3 \times 10^8 (q/q_e)BR, \tag{34.9}$$

donde B, R y la velocidad de la luz están todas expresadas en el sistema mks, siendo la última numéricamente igual a 3×10^8.

La unidad mks de campo magnético se llama *wéber por metro cuadrado*. Hay una unidad más antigua que está aún en uso, llamada *gauss*. Un wéber/m^2 es igual a 10^4 gauss. Para dar una idea de lo grandes que son los campos magnéticos, el campo magnético más intenso que se puede producir en el hierro es de 1.5×10^4 gauss; más allá de eso, la ventaja de usar hierro desaparece. Hoy día los electroimanes con bobinas superconductoras son capaces de producir campos

constantes de intensidad superior a 10^5 gauss —esto es, 10 unidades mks—. El campo de la Tierra en el ecuador es de unas pocas décimas de gauss.

Volviendo a la ecuación (34.9) podríamos imaginar el sincrotrón funcionando con 1 000 millones de electronvoltios, así pc sería 10^9 para 1 000 millones de electronvoltios. Volveremos a la energía dentro de un momento. Luego, si tuviéramos un B correspondiente a, digamos, 10 000 gauss, que es un campo bastante considerable, una unidad mks, entonces vemos que R debería ser 3.3 metros. El radio real del sincrotrón del Caltech es de 3.7 metros, el campo es un poco más grande y la energía es de 1 500 millones, pero la idea es la misma. Así, pues, tenemos ahora una idea de por qué el sincrotrón tiene el tamaño que tiene.

Hemos calculado el momentum, pero sabemos que la energía total, incluyendo la energía en reposo, está dada por $W = \sqrt{p^2c^2 + m^2c^4}$, y para un electrón la energía en reposo correspondiente a mc^2 es de 0.511×10^6 eV, de modo que cuando pc es 10^9 eV podemos omitir mc^2, y así, para todos los fines prácticos, $W = pc$ cuando las velocidades son relativistas. Es prácticamente lo mismo decir que la energía de un electrón es 1 000 millones de electronvoltios, que decir que el momentum por c es 1 000 millones de electronvoltios. Si $W = 10^9$ eV ¡es fácil demostrar que la velocidad difiere de la velocidad de la luz en sólo una parte en ocho millones!

Volvamos ahora a la radiación emitida por tal partícula. Una partícula moviéndose en un círculo de radio 3.3 metros, o de 20 metros de circunferencia, da una vuelta en aproximadamente el tiempo que demora la luz en viajar 20 metros. Entonces la longitud de onda que esa partícula debería emitir debería ser de 20 metros —en la región de onda corta de radio—. Pero debido al efecto de amontonamiento que hemos estado discutiendo (figura 34-3) y debido a que la distancia en la cual debemos extender el radio para alcanzar la velocidad c es solamente de una parte en ocho millones del radio, las cúspides de las hipocicloides son enormemente agudas comparadas con la distancia entre ellas. La aceleración, que implica una segunda derivada respecto al tiempo, obtiene el doble del "factor de compresión" de 8×10^6 debido a que la escala de tiempo se reduce en ocho millones dos veces en las vecindades de la cúspide. Así, pues, debemos esperar que la longitud de onda efectiva sea mucho más corta, del orden de 64 veces 10^{12} más pequeña que 20 metros, y eso corresponde a la región de rayos X. (Actualmente, la cúspide misma no es el único factor determinante, debe también incluirse una cierta región alrededor de la cúspide. Esto cambia el factor a la potencia 3/2 en vez del cuadrado, pero aún nos deja por encima de la región óptica.) Así, aun cuando un electrón moviéndose lentamente habría irradiado ondas de 20 metros, ¡el efecto relativista acorta la longitud de onda tanto que la podemos *ver*! Evidentemente, la luz debería estar *plarizada* con el vector eléctrico perpendicular al campo magnético uniforme.

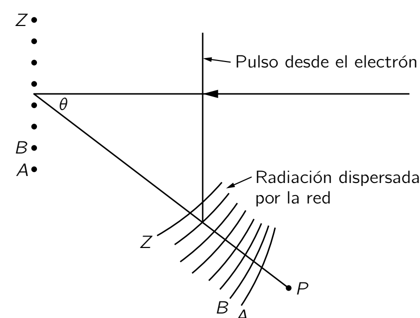


Figura 34-5. La luz, que incide sobre una red como un simple pulso agudo, se dispersa en diversas direcciones como luz de diferentes colores.

Para apreciar mejor qué observaríamos, supongan que fuéramos a tomar esa luz (para simplificar las cosas, dado que estos pulsos están tan separados en el tiempo, tomaremos solamente un pulso) y la dirigimos sobre una red de difracción, que consiste en muchos alambres dispersores. Después que este pulso se aleja de la red, ¿qué vemos? (Deberíamos ver luz roja, luz azul, etc., si es que vemos alguna luz.) ¿Qué vemos *efectivamente*? El pulso golpea la red centralmente, y todos los osciladores de la red, juntos, son violentamente movidos hacia arriba y luego hacia abajo de nuevo, solamente una vez. Producen entonces efectos en varias direcciones, como muestra la figura 34-5. Pero el punto P está más cerca de un extremo del retículo que del otro, por lo que en este punto el campo eléctrico llega primero desde el alambre A, después desde B, y así sucesivamente; finalmente, llega el pulso del último alambre. En resumen, la suma de las reflexiones de todos los alambres sucesivos se muestra en la figura 34-6(a); es un campo eléctrico que es una serie de pulsos, y es muy parecido a una onda sinusoidal, cuya longitud de onda es la distancia entre los pulsos, ¡justamente como lo sería para luz monocromática incidiendo sobre la red! Así, obtenemos luz coloreada, de veras. Pero, con el mismo razonamiento, ¿no obtendremos luz a partir de cualquier tipo de "pulso"? No. Supongan que la curva fuera mucho más suave; entonces sumaríamos todas las ondas dispersadas, separadas por un pequeño tiempo entre

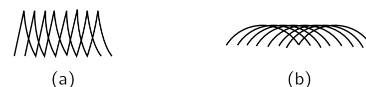


Figura 34-6. El campo eléctrico total debido a una serie de (a) pulsos agudos y (b) pulsos suaves.

ellas (figura 34-6b). Luego vemos que el campo no se sacudiría en absoluto, sería una curva muy suave porque cada pulso no varía mucho en el intervalo de tiempo entre los pulsos.

La radiación electromagnética emitida por partículas relativistas cargadas circulando en un campo magnético, se llama *radiación sincrotrónica cósmica*. Se llama así por razones obvias, pero no está específicamente limitada a los sincrotrones, o siquiera a laboratorios terrestres. ¡Es excitante e interesante que también ocurre en la naturaleza!



(a)



(b)

Figura 34-7. La nebulosa del cangrejo vista a través de un filtro azul y un polaroide. (a) El vector eléctrico vertical. (b) El vector eléctrico horizontal.

34-4 Radiación sincrotrónica cósmica

En el año 1054, las civilizaciones china y japonesa se encontraban entre las más avanzadas del mundo; estaban conscientes del universo externo, y registraron aquel año algo muy notable: una estrella explosiva y brillante. (Es extraordinario que ninguno de los monjes europeos, quienes escribían todos los libros en la Edad Media, ni siquiera se molestaran en escribir que explotó una estrella en el cielo.) Hoy podemos tomar una fotografía de aquella estrella, y lo que vemos se muestra en la figura 34-8. En el exterior hay una gran masa de filamentos rojos, que se produce por los átomos del gas diluido "sonando" con sus frecuencias naturales; esto forma un brillante espectro de líneas con diferentes frecuencias en él. Resulta que, en este caso, el rojo se debe al nitrógeno. En cambio, en la región central hay una misteriosa mancha de luz borrosa con una distribución *continua* de frecuencias, es decir, no hay frecuencias especiales asociadas con átomos individuales. Tampoco es polvo "encendido" por las estrellas vecinas, lo que es una manera de obtener un espectro continuo. Podemos ver estrellas a través de ella; así que es transparente, pero está *emitiendo* luz.

En la figura 34-7 observamos el mismo objeto, usando luz en una región del espectro que no tiene líneas espectrales brillantes, así que podemos ver solamente la región central. Pero en este caso también se han puesto polarizadores en el telescopio, y las dos vistas corresponden a dos orientaciones, diferentes en 90°. ¡Vemos que las fotografías son distintas! Es decir, la luz está polarizada. La razón es, presumiblemente, que hay un campo magnético local, y muchos electrones muy rápidos moviéndose en aquel campo magnético.

Hemos ilustrado ya cómo los electrones pueden dar vueltas alrededor del campo en un círculo. A esto podemos agregar, naturalmente, cualquier movimiento uniforme en la dirección del campo, puesto que la fuerza $q\boldsymbol{v} \times \boldsymbol{B}$ no tiene



Figura 34-8. La nebulosa del cangrejo vista con todos los colores (sin filtro).

componente en esta dirección, y como hemos indicado, la radiación sincrotrónica está evidentemente polarizada en dirección perpendicular a la proyección del campo magnético sobre el plano visual.

Reuniendo estos dos hechos, vemos que en una región donde una fotografía es brillante y en la otra es negra, la luz debe tener su campo eléctrico completamente polarizado en una dirección. Esto significa que hay un campo magnético perpendicular a esta dirección, mientras en las otras regiones, donde hay una fuerte emisión en la otra fotografía, el campo magnético debe estar de la otra manera. Si observamos cuidadosamente la figura 34-7, podemos notar que hay, en términos aproximados, un conjunto general de “líneas” que van en una dirección en una foto, y perpendicularmente a ésta en la otra. Las fotografías muestran un tipo de estructura fibrosa. Presumiblemente, las líneas del campo magnético tenderán a extenderse a distancias relativamente grandes en su propia dirección, y así, presumiblemente, hay grandes regiones de campo magnético con todos los electrones dando vueltas en espiral en un sentido, mientras en otra región el campo está en otro sentido, y los electrones están también dando vueltas en espiral en ese sentido.

¿Qué es lo que mantiene tan alta la energía de los electrones durante tanto tiempo? Después de todo ya hace 900 años que se produjo la explosión. ¿Cómo pueden seguir moviéndose tan rápido? Cómo mantienen su energía y cómo toda esta cosa sigue evolucionando no se entiende todavía completamente.

34-5 Bremsstrahlung

A continuación mencionaremos brevemente otro interesante efecto de una partícula en movimiento rápido que irradia energía. La idea es muy similar a la que ya hemos discutido. Suponga que en un pedazo de materia hay partículas cargadas y un electrón muy rápido, digamos, se aproxima (figura 34-9). Entonces, debido al campo eléctrico alrededor del núcleo atómico, el electrón es atraído, acelerado, de manera que la curva de su movimiento tiene un ligero codo o dobladura en ella. Si el electrón está viajando casi a la velocidad de la luz, ¿cuál será el campo eléctrico producido en dirección C? Recuerden nuestra regla: tomamos el movimiento real, lo trasladamos hacia atrás con la velocidad c, y eso da una curva cuya curvatura mide el campo eléctrico. Estaba aproximándose hacia nosotros con la velocidad v, así obtenemos un movimiento hacia atrás con el cuadro completo comprimido dentro de una distancia menor en la proporción en que $c - v$ es menor que c. Así, si $1 - v/c \ll 1$, hay una curvatura muy aguda y rápida en B', y cuando tomamos la segunda derivada de aquélla, obtenemos un campo muy fuerte en la dirección del movimiento. Así, cuando electrones muy rápidos se mueven a través de la materia, lanzan radiación en la dirección hacia adelante. Esto se llama *de Bremsstrahlung*. En realidad, el sincrotrón se usa, no tanto para producir electrones de alta energía (realmente si pudiéramos sacarlos de la máquina más convenientemente, no diríamos esto), sino para producir fotones de energía muy alta (rayos gamma), haciendo pasar los electrones a través de un “blanco” de tungsteno sólido, y dejándolos irradiar fotones mediante este efecto de bremsstrahlung.

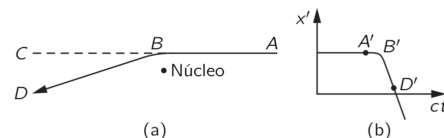


Figura 34-9. Un electrón rápido al pasar cerca de un núcleo irradia energía en la dirección de su movimiento.

34-6 El efecto Doppler

Pasemos ahora a considerar algunos otros efectos de fuentes en movimiento. Supongan que la *fuente* es un átomo quieto que está oscilando a una de sus frecuencias naturales, ω_0. Entonces sabemos que la frecuencia de la luz que observaríamos es ω_0. Pero ahora tomemos otro ejemplo, en el cual tenemos un oscilador similar oscilando con una frecuencia ω_1, y al mismo tiempo el átomo entero, todo oscilador, se mueve con velocidad v y en dirección hacia el observador. Entonces, el movimiento real en el espacio es, naturalmente, como se muestra en la figura 34-10(a). Ahora hacemos nuestro juego usual, sumamos $c\tau$; es decir, trasladamos la curva entera hacia atrás y encontramos entonces que oscila como en la figura 34-10(b). En una cantidad dada de tiempo τ, cuando el oscilador

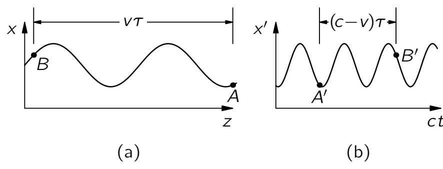


Figura 34-10. Las curvas *x-z* y *x′-t* de un oscilador en movimiento.

debería haber viajado una distancia $v\tau$, sobre el diagrama x' en función de ct viaja una distancia $(c-v)\tau$. Así pues, todas las oscilaciones de frecuencia ω_1 en el tiempo $\Delta\tau$ se encuentran ahora en el intervalo $\Delta t = (1-v/c)\Delta\tau$; se aplastan y a medida que esta curva pase por nosotros a velocidad c, veremos luz de una *frecuencia más alta*, más alta en justamente el factor de compresión $(1-v/c)$. Entonces observamos

$$\omega = \frac{\omega_1}{1-v/c}. \tag{34.10}$$

Naturalmente, podemos analizar esta situación de varias maneras. Supongan que el átomo estuviera emitiendo, en vez de ondas sinusoidales, una serie de pulsos, pi, pi, pi, pi, con cierta frecuencia ω_1. ¿Con qué frecuencia los recibiríamos? El primero llega con cierto retardo, pero el próximo está menos retardado porque mientras tanto el átomo se mueve más cerca del receptor. Por lo tanto, el tiempo entre los “pi” disminuye por el movimiento. Si analizamos la geometría de la situación, encontramos que la frecuencia de los pi aumenta en el factor $1/(1-v/c)$.

¿Será entonces $\omega = \omega_0/(1-v/c)$ la frecuencia que se observaría si tomáramos un átomo ordinario, que tuviera una frecuencia natural ω_0, y se moviera con velocidad v hacia el observador? No, sabemos la frecuencia natural ω_1 de un átomo en movimiento no es la misma que la medida cuando está quieto, debido a la dilatación relativista de la velocidad del transcurrir del tiempo. Entonces, si ω_0 fuera la frecuencia natural modificada, ω_1 sería

$$\omega_1 = \omega_0\sqrt{1-v^2/c^2}. \tag{34.11}$$

Por lo tanto, la frecuencia observada ω es

$$\omega = \frac{\omega_0\sqrt{1-v^2/c^2}}{1-v/c}. \tag{34.12}$$

El corrimiento de frecuencia observado en la situación anterior se llama *efecto Doppler*: si algo se mueve hacia nosotros, la luz que emite aparece más violeta, y si se aleja aparece más roja.

Daremos ahora dos deducciones más de este mismo interesante e importante resultado. Supongan ahora que la *fuente* está quieta y está emitiendo ondas de frecuencia ω_0, mientras que el *observador* está moviéndose hacia la fuente a la velocidad v. Después de cierto periodo de tiempo t, el observador se habrá movido a una nueva posición, una distancia vt desde donde él estaba en $t = 0$. ¿Cuántos radianes de fase habrá visto pasar? Un cierto número, $\omega_0 t$, pasaron por cualquier punto fijo, y además el observador ha barrido en su pasada algunos más debido a su propio movimiento, precisamente un número de vtk_0 (el número de radianes por metro multiplicado por la distancia). Así, el número total de radianes en el tiempo t, o sea la frecuencia observada, sería $\omega_1 = \omega_0 + k_0 v$. Hemos hecho este análisis desde el punto de vista de un hombre en reposo; nos gustaría saber qué le parecería al hombre que está moviéndose. Aquí tenemos que preocuparnos de nuevo acerca de la diferencia en la rapidez de marcha del reloj para los dos observadores, y esta vez eso significa que tenemos que *dividir* enttre $\sqrt{1-v^2/c^2}$. Así, si k_0 es el número de onda, el número de radianes por metro en la dirección del movimiento, y ω_0 es la frecuencia, entonces la frecuencia observada por el hombre en movimiento es:

$$\omega = \frac{\omega_0 + k_0 v}{\sqrt{1-v^2/c^2}}. \tag{34.13}$$

Para el caso de la luz, sabemos que $k_0 = \omega_0/c$. Así que en este problema particular, la ecuación se leería

$$\omega = \frac{\omega_0(1+v/c)}{\sqrt{1-v^2/c^2}}, \tag{34.14}$$

¡que parece completamente diferente a la fórmula (34.12)! La frecuencia que observaríamos si nos movemos hacia una fuente ¿será diferente de la frecuencia

que veríamos si la fuente se moviera hacia nosotros? ¡Naturalmente no! La teoría de la relatividad dice que las dos deben ser *exactamente iguales*. ¡Si fuéramos matemáticos suficientemente expertos probablemente nos daríamos cuenta de que esas dos expresiones matemáticas son exactamente iguales! De hecho, la igualdad *necesaria* de las dos expresiones es una de las formas con que a algunas personas les gusta demostrar que la relatividad exige una dilatación del tiempo, porque si no hubiéramos puesto aquellos factores de raíz cuadrada, no seguirían siendo iguales.

Como sabemos relatividad, analicémoslo de otra manera, que puede parecer un poco más general. (¡Es realmente la misma cosa, puesto que no importa *cómo* lo hacemos!) De acuerdo con la teoría de la relatividad, hay una relación entre la posición y el tiempo observado por un hombre, y la posición y el tiempo visto por otro que está moviéndose respecto a él. Escribimos esas relaciones tiempo atrás (capítulo 16). Es la *transformación de Lorentz* y su inversa:

$$\begin{aligned} x' &= \frac{x+vt}{\sqrt{1-v^2/c^2}}, \qquad & x &= \frac{x'-vt'}{\sqrt{1-v^2/c^2}}, \\ t' &= \frac{t+vx/c^2}{\sqrt{1-v^2/c^2}}, \qquad & t &= \frac{t'-vx'/c^2}{\sqrt{1-v^2/c^2}}. \end{aligned} \tag{34.15}$$

Si estuviéramos inmóviles en el suelo, la forma de una onda sería $\cos(\omega t - kx)$; todos los nodos y máximos y mínimos seguirían esta forma. Pero, ¿qué vería un hombre en movimiento, observando la misma onda física? Donde el campo es cero, las posiciones de todos los nodos son las mismas (cuando el campo es *cero*, *todos* miden el campo como cero); es un invariante relativista. Así, pues, la forma es la misma también para el otro hombre, excepto que debemos transformarla a su sistema de referencia:

$$\cos(\omega t - kx) = \cos\left[\omega\,\frac{t'-vx'/c^2}{\sqrt{1-v^2/c^2}} - k\,\frac{x'-vt'}{\sqrt{1-v^2/c^2}}\right].$$

Si reagrupamos los términos dentro de los paréntesis, obtenemos

$$\begin{aligned} \cos(\omega t - kx) &= \cos\Bigg[\underbrace{\frac{\omega+kv}{\sqrt{1-v^2/c^2}}}\,t' - \underbrace{\frac{k+v\omega/c^2}{\sqrt{1-v^2/c^2}}}\,x'\Bigg] \\ &= \cos\,[\qquad \omega' \qquad t' - \qquad k' \qquad x'\,]. \end{aligned} \tag{34.16}$$

Ésta es una onda de nuevo, una onda coseno, en la cual hay una cierta frecuencia ω', una constante multiplicando a t' y otra constante, k', que multiplica a x'. Llamamos k' el número de onda, o número de ondas por metro, para el otro hombre. Por lo tanto, el otro hombre verá una nueva frecuencia y un nuevo número de onda dado por

$$\omega' = \frac{\omega+kv}{\sqrt{1-v^2/c^2}}, \tag{34.17}$$

$$k' = \frac{k+\omega v/c^2}{\sqrt{1-v^2/c^2}}. \tag{34.18}$$

Si observamos (34.17), vemos que es la misma fórmula (34.13) que obtuvimos con un razonamiento más físico.

34-7 El vector de dimensión cuatro $\boldsymbol{\omega}, \boldsymbol{k}$

Las relaciones indicadas en las ecuaciones (34.17) y (34.18) son muy interesantes, porque dicen que la nueva frecuencia ω' es una combinación de la antigua frecuencia ω y el antiguo número de onda k, y que el nuevo número de onda es

una combinación de los antiguos números de onda y frecuencia. Ahora bien, el número de onda es la velocidad de cambio de fase con la distancia, y la frecuencia es la velocidad de cambio de fase con el tiempo, y en esas expresiones vemos una estrecha analogía con la transformación de Lorentz de la posición y del tiempo: si se piensa que ω es como t, y si se piensa que k es como x dividido entre c', el nuevo ω' sería como t' y el nuevo k' sería como x'/c^2. Es decir, *frente a la transformación de Lorentz, ω y k se transforman en la misma forma que t y x*. Ellos constituyen lo que llamamos un *vector de dimensión cuatro*; cuando una cantidad tiene cuatro componentes que se transforman como tiempo y espacio, es un vector de dimensión cuatro. Todo parece bien, entonces, excepto en una pequeña cosa: dijimos que un vector de dimensión cuatro tiene que tener *cuatro componentes*; ¿dónde están las otras dos componentes? Hemos visto que ω y k son como tiempo y espacio en una dirección espacial, pero no en todas direcciones, y así debemos estudiar a continuación el problema de la propagación de la luz en tres dimensiones espaciales, no sólo en una dirección, como lo hemos estado haciendo hasta ahora.

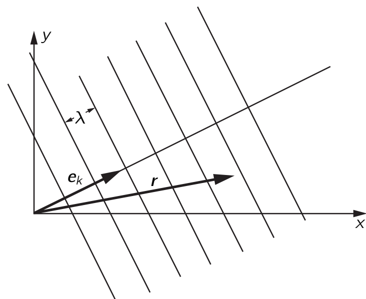


Figura 34-11. Una onda plana que viaja en una dirección oblicua.

Supongan que tenemos un sistema de coordenadas, x, y, z, y una onda que se propaga y cuyos frentes de ondas son como se muestra en la figura 34-11. La longitud de onda de la onda es λ, pero la dirección de movimiento de la onda no está en la dirección de uno de los ejes. ¿Cuál es la fórmula para esa onda? La respuesta es claramente $\cos(\omega t - ks)$, donde $k = 2\pi/\lambda$ y s es la distancia según la dirección de movimiento de la onda —la componente de la posición espacial en la dirección de movimiento—. Pongámoslo de esta manera: si $\boldsymbol{r}$ es el vector posición de un punto en el espacio, s es $\boldsymbol{r} \cdot \boldsymbol{e}_k$, donde $\boldsymbol{e}_k$ es un vector unitario en la dirección de movimiento. O sea, s es justamente $r\cos(\boldsymbol{r} \cdot \boldsymbol{e}_k)$, la componente de la distancia en la dirección de movimiento. Por lo tanto, nuestra onda es $\cos(\omega t - k\boldsymbol{e}_k \cdot \boldsymbol{r})$.

Resulta ahora conveniente definir un vector $\boldsymbol{k}$, llamado *vector de onda*, que tiene un módulo igual al número de onda, $2\pi/\lambda$, y está dirigido según la dirección de propagación de las ondas:

$$\boldsymbol{k} = 2\pi\boldsymbol{e}_k/\lambda = k\boldsymbol{e}_k. \tag{34.19}$$

Usando este vector, nuestra onda puede escribirse como $\cos(\omega t - \boldsymbol{k} \cdot \boldsymbol{r})$ o como $\cos(\omega t - k_x x - k_y y - k_z z)$. ¿Cuál es el significado de una componente de $\boldsymbol{k}$, digamos k_x? Evidentemente, k_x es la velocidad de desfase respecto a x. Refiriéndonos a la figura 34-11, vemos que la fase varía cuando variamos x, como si hubiera una onda a lo largo de x, *pero de una longitud de onda más larga*. La "longitud de onda en la dirección x" es más larga que una longitud de onda natural verdadera, en la secante del ángulo α entre la dirección real de propagación y el eje x:

$$\lambda_x = \lambda/\cos\alpha. \tag{34.20}$$

Por lo tanto, la velocidad de cambio de fase, que es proporcional a la *inversa* de λ_x, es *más pequeña* por el factor $\cos\alpha$; así es justamente como variaría k_x —¡sería el módulo de $\boldsymbol{k}$ multiplicado por el coseno del ángulo entre $\boldsymbol{k}$ y el eje x!—

Ésa es, entonces, la naturaleza del vector de onda que usamos para representar una onda en tres dimensiones. Las cuatro cantidades ω, k_x, k_y, y k_z se transforman en relatividad en un vector de dimensión cuatro, donde ω corresponde al tiempo, y $k_x, k_y, \text{y} k_z$ corresponden a las componentes x, y y z del vector de dimensión cuatro.

En nuestra discusión anterior de la relatividad especial (capítulo 17) aprendimos que hay formas de hacer productos escalares relativistas con vectores de dimensión cuatro. Si usamos el vector posición $\boldsymbol{x}_\mu$, donde μ representa las cuatro componentes (tiempo y tres espaciales), y si llamamos vector de onda $k_\mu = (\omega, \boldsymbol{k})$, donde el índice μ de nuevo tiene cuatro valores, tiempo y tres espaciales, luego el producto escalar de x_μ y k_μ se escribe $\sum' k_\mu x_\mu$ (véase capítulo 17). Este producto escalar es un invariante, independiente del sistema de coordenadas; ¿a qué es igual? Por la definición de este producto escalar en cuatro dimensiones, es

$$\sum{}' k_\mu x_\mu = \omega t - k_x x - k_y y - k_z z. \tag{34.21}$$

Sabemos por nuestro estudio de vectores que $\sum' k_\mu x_\mu$ es invariante frente a la transformación de Lorentz, puesto que k_μ es un vector de dimensión cuatro. Pero esta cantidad es precisamente lo que aparece dentro del coseno para una onda plana, y *debería* ser invariante frente a una transformación de Lorentz. No podemos tener una fórmula con algo que cambia en el interior del coseno, puesto que sabemos que la fase de una onda no puede cambiar cuando cambiamos el sistema de coordenadas.

34-8 Aberración

Al derivar las ecuaciones (34.17) y (34.18) hemos tomado un simple ejemplo donde $\boldsymbol{k}$ resultó estar en la dirección de movimiento, pero naturalmente podemos generalizarlo también a otros casos. Por ejemplo, supongan que hay una fuente emitiendo luz en cierta dirección desde el punto de vista de un hombre en reposo, pero nos estamos moviendo sobre la Tierra, digamos (figura 34-12). ¿De qué dirección parece venir la luz? Para averiguarlo tendremos que escribir las cuatro componentes de k_μ y aplicar la transformación de Lorentz. La respuesta, sin embargo, puede encontrarse con el siguiente razonamiento: tenemos que apuntar nuestro telescopio en un ángulo para ver la luz. ¿Por qué? Porque la luz se aproxima con la velocidad c y nosotros nos estamos moviendo de lado con la velocidad v, y así que hay que inclinar el telescopio hacia delante de tal manera que cuando la luz llegue, pase "derecho" por el tubo. Es muy fácil ver que la distancia horizontal es vt cuando la distancia vertical es ct, y por lo tanto, si θ' es el ángulo de inclinación, $\tan\theta' = v/c$. ¡Qué bonito! Qué bonito, realmente excepto en una pequeña cosa: θ' *no* es el ángulo en que se debe colocar el telescopio, *respecto a la Tierra*, porque hemos hecho nuestro análisis desde el punto de vista de un observador "fijo". Cuando dijimos que la distancia horizontal es vt, el hombre sobre la Tierra habría encontrado una distancia diferente, puesto que él ha medido con una regla "aplastada". Resulta que, debido al efecto de esa contracción

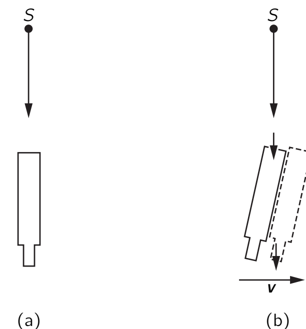


Figura 34-12. Una fuente S distante se observa con (a) un telescopio quieto y (b) un telescopio que se mueve hacia un lado.

$$\tan\theta = \frac{v/c}{\sqrt{1 - v^2/c^2}}, \tag{34.22}$$

que es equivalente a

$$\operatorname{sen}\theta = v/c. \tag{34.23}$$

Será instructivo para el estudiante deducir este resultado, usando la transformación de Lorentz.

Este efecto, que un telescopio tiene que ser inclinado, se llama *aberración*, y ha sido observado. ¿*Cómo* podemos observarlo? ¿Quién puede decir dónde *debería* estar una estrella dada? Supongamos que *tenemos* que mirar en la dirección equivocada para ver una estrella; ¿cómo sabemos que es una dirección equivocada? Porque la Tierra da vueltas alrededor del Sol. Hoy tenemos que apuntar el telescopio de una manera; seis meses más tarde tenemos que inclinar el telescopio de otra manera. Así es como podemos decir que existe tal efecto.

34-9 El momentum de la luz

Cambiemos ahora a un tópico diferente. Nunca hemos dicho, en todas nuestras discusiones de los pocos capítulos pasados, cosa alguna acerca de los efectos del campo *magnético* que está asociado con la luz. Ordinariamente, los efectos del campo magnético son muy pequeños, pero hay un efecto interesante e importante, que es una consecuencia del campo magnético. Supongan que viene luz de una fuente y está actuando sobre una carga, y forzándola de un lado a otro. Supondremos que el campo eléctrico está en la dirección x, así que el movimiento de la carga está también en la dirección x: tiene una posición x y una velocidad

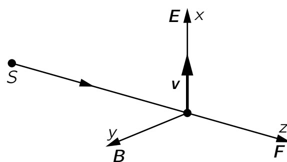


Figura 34-13. La fuerza magnética sobre una carga impulsada por un campo eléctrico está en la dirección del rayo de luz.

v, como muestra la figura 34-13. El campo magnético es perpendicular al campo eléctrico. Ahora bien, a medida que el campo eléctrico actúa sobre la carga, y la mueve hacia arriba y hacia abajo, ¿qué hace el campo magnético? El campo magnético actúa sobre la carga (digamos un electrón) solamente cuando ésta se mueve; pero el electrón *está* moviéndose, está forzado por el campo eléctrico, así que ambos trabajan juntos: mientras la cosa viaja hacia arriba y hacia abajo, tiene una velocidad y hay una fuerza sobre ella, B por v por q; pero, ¿en qué *dirección* está esta fuerza? *Está en la dirección de propagación de la luz.* Por lo tanto, cuando la luz está iluminando una carga, y oscila en respuesta a esa luz, hay una fuerza impulsora en la dirección del rayo de luz. Esto se llama *presión de radiación* o presión de luz.

Determinemos cuánto vale la presión de radiación. Evidentemente ésta es $F = qvB$, o puesto que todo está oscilando, es el *promedio en el tiempo* de esto, $\langle F\rangle$. Según la ecuación (34.2) la intensidad del campo magnético es igual a la intensidad del campo eléctrico dividida entre c, así que necesitamos encontrar el promedio del campo eléctrico, por la velocidad, por la carga, por $1/c$: $\langle F\rangle = q\langle vE\rangle/c$. ¡Pero la carga q por el campo E es la fuerza eléctrica sobre la carga, y la fuerza sobre la carga por la velocidad es el trabajo dW/dt realizado sobre la carga! ¡Por lo tanto, la fuerza, el "momentum impulsor" que se entrega por segundo por la luz, es igual a $1/c$ por la *energía absorbida* de la luz por segundo! Ésa es una regla general, puesto que no dijimos cuál era la intensidad de la radiación, o si algunas de las cargas se anulaban. *En cualquier circunstancia donde la luz está siendo absorbida, hay una presión.* El momentum que entrega la luz es siempre igual a la energía absorbida, dividida entre c:

$$\langle F\rangle = \frac{dW/dt}{c}. \tag{34.24}$$

Ya sabemos que la luz transporta energía. Ahora comprendemos que también transporta *momentum* y, además, que el momentum transportado es siempre igual a $1/c$ veces la energía.

Cuando una fuente emite luz hay un efecto de retroceso: la misma cosa a la inversa. Si un átomo está emitiendo una energía W en alguna dirección, entonces hay un momentum de retroceso $p = W/c$. Si se *refleja* luz perpendicularmente en un espejo, obtenemos dos veces la fuerza.

Hasta aquí llegamos usando la teoría clásica de la luz. Por cierto que sabemos que hay una teoría cuántica, y que en muchos aspectos la luz actúa como una partícula. La energía de una partícula de luz es una constante multiplicada por la frecuencia:

$$W = h\nu = \hbar\omega. \tag{34.25}$$

Ahora notamos que la luz también transporta un momentum igual a la energía dividida entre c, así que también es verdad que esas partículas efectivas, estos *fotones*, llevan momentum,

$$p = W/c = \hbar\omega/c = \hbar k. \tag{34.26}$$

La *dirección* del momentum es, naturalmente, la dirección de propagación de la luz. Así, para ponerlo en forma vectorial,

$$W = \hbar\omega, \qquad \boldsymbol{p} = \hbar\boldsymbol{k}. \tag{34.27}$$

Sabemos también, naturalmente, que la energía y el momentum de una partícula deberían formar un vector de dimensión cuatro. Hemos descubierto recién que ω y $\boldsymbol{k}$ forman un vector de dimensión cuatro. Por lo tanto, es bueno que (34.27) tenga la misma constante en ambos casos; significa que la teoría cuántica y la teoría de la relatividad son mutuamente compatibles.

La ecuación (34.27) se puede escribir más elegantemente como $p_\mu = \hbar k_\mu$, una ecuación relativista, para una partícula asociada con una onda. Aunque hemos discutido esto sólo para fotones, para los cuales k (el módulo de $\boldsymbol{k}$) es igual a

ω/c y $p = W/c$, la relación es mucho más general. En mecánica cuántica *todas* las partículas, no solamente los fotones, muestran propiedades ondulatorias, pero la frecuencia y el número de onda de las ondas está relacionado con la energía, y el momentum de las partículas por (34.27) (que se llaman relaciones de Broglie), aun cuando p no es igual a W/c.

En el último capítulo vimos que un rayo de luz polarizada circularmente, derecha o izquierda, también lleva *momentum angular* en una cantidad proporcional a la energía $\mathcal{E}$ de la onda. En la representación cuántica, un rayo de luz circularmente polarizada se considera como un flujo de fotones, cada uno de los cuales lleva un momentum angular $\pm\hbar$ según la dirección de propagación. Eso es lo que llega a ser la polarización desde el punto de vista corpuscular: los fotones llevan momentum angular como balas de rifle girando sobre sí mismas. Pero esta imagen "balística" es realmente tan incompleta como la imagen "ondulatoria", y tendremos que discutir esas ideas más exhaustivamente en un capítulo posterior sobre comportamiento cuántico.

35

Visión de los colores

35-1 El ojo humano

El fenómeno de los colores depende parcialmente del mundo físico. Discutimos los colores de películas de jabón, y de otras cosas, como producidos por interferencia. Pero también, por cierto, depende del ojo, o lo que sucede detrás del ojo, dentro del cerebro. La física caracteriza la luz que entra al ojo, sin embargo, nuestras sensaciones son el resultado de procesos neurofotoquímicos y de respuestas psicológicas.

Existen muchos fenómenos interesantes asociados con la visión que comprenden una mezcla de fenómenos físicos y procesos fisiológicos, y la total apreciación de los fenómenos naturales; cuando nosotros los vemos, debe ir más allá de la física en el sentido usual. No damos mayores justificaciones para hacer estas excursiones a otros campos, porque la separación de campos, como lo hemos recalcado, es meramente una conveniencia humana, y no una cosa natural. La naturaleza no está interesada en nuestras divisiones, y muchos fenómenos interesantes tienden un puente sobre las brechas entre campos.

En el capítulo 3 ya hemos discutido la relación entre la física y las otras ciencias en términos generales, pero ahora vamos a mirar con algún detalle un campo específico en el cual la física y otras ciencias están muy, pero muy relacionadas. Tal área es la *visión*. En particular discutiremos la *visión de los colores*. En el capítulo presente discutiremos principalmente los fenómenos observables de la visión humana, y en el capítulo siguiente consideraremos los aspectos fisiológicos de la visión, tanto en el hombre como en otros animales.

Todo comienza con el ojo; así, con el fin de entender cuál fenómeno vemos, se requiere cierto conocimiento del ojo. En el capítulo siguiente discutiremos con algún detalle cómo trabajan las diferentes partes del ojo, y cómo están interconectadas con el sistema nervioso. Por ahora, describiremos sólo brevemente cómo funciona el ojo (figura 35-1).

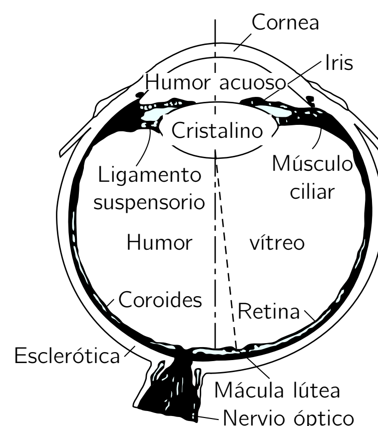


Figura 35-1. El ojo.

La luz entra al ojo a través de la *córnea*; ya hemos discutido cómo se desvía para formar imagen en una capa llamada *retina* en la parte posterior del ojo, de manera que diferentes partes de la retina reciben luz desde distintas partes del campo visual exterior. La retina no es absolutamente uniforme: existe un lugar, un punto, en el centro de nuestro campo visual, que utilizamos cuando tratamos de ver muy cuidadosamente las cosas, y en el cual tenemos la visión más aguda, se llama *fóvea* o *mácula*. Las partes laterales del ojo, como podemos notar inmediatamente a partir de nuestra experiencia al mirar las cosas, no son tan efectivas para ver detalles, como lo es el centro del ojo. Existe también, un punto en la retina de donde salen los nervios que llevan la información, éste es un punto ciego. No existe aquí parte sensible de la retina, y es posible demostrar que si cerramos, digamos, el ojo izquierdo y miramos directamente un objeto, y en seguida movemos un dedo u otro objeto pequeño lentamente hacia fuera del campo visual, súbitamente desaparece en alguna parte. El único uso práctico que de este hecho conocemos es que cierto fisiólogo llegó a ser todo un favorito en la corte del rey de Francia, a quien se lo dio a conocer; en las aburridas sesiones con sus cortesanos, el rey podía divertirse "cortándoles la cabeza", mirando a alguno y viendo desaparecer la cabeza del otro.

La figura 35-2 muestra una vista ampliada del interior de la retina en forma algo esquemática. En diferentes partes de la retina existen distintas clases de estructuras. Los objetos que se presentan más densamente cerca de la periferia de la retina se llaman *bastoncitos*. Más cerca de la fóvea encontramos, además

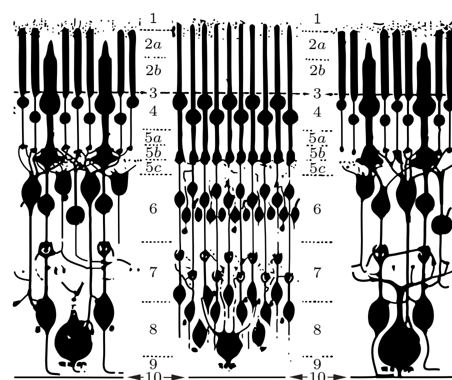


Figura 35-2. La estructura de la retina (la luz entra desde abajo).

de estas células bastoncitos, las células *conos*. Describiremos más adelante estas células. A medida que nos aproximamos a la fóvea, aumenta el número de conos, y en la fóvea misma no hay otra cosa que células conos, empaquetados en forma muy compacta, tan compacta que las células conos son aquí mucho más finas o delgadas que en cualquier otra parte. Debemos así apreciar que vemos con los conos justamente en el centro del campo visual, pero a medida que nos vamos a la periferia tenemos las otras células, los bastoncitos. Ahora la cosa interesante es que en la retina cada célula sensible a la luz no está conectada por una fibra directa al nervio óptico, sino que está conectada a muchas otras células, que a su vez están conectadas entre sí. Existen varias clases de células: hay células que llevan la información hasta el nervio óptico, pero hay otras que están principalmente interconectadas "horizontalmente". Existen esencialmente cuatro clases de células, pero no entraremos en estos detalles ahora. Lo principal en que ponemos énfasis es que la señal luminosa ha sido ya "pensada". Esto quiere decir que la información desde las diferentes células no va inmediatamente al cerebro, punto por punto, sino que en la retina una cierta cantidad de la información ya ha sido digerida por una combinación de la información desde diversos receptores visuales. Es importante comprender que ciertos fenómenos de la función cerebral ocurren en el ojo mismo.

35-2 El color depende de la intensidad

Uno de los fenómenos más sorprendentes de la visión es la adaptación del ojo a la oscuridad. Si nos introducimos en la oscuridad desde una pieza brillantemente iluminada, no podemos ver muy bien durante un instante, pero paulatinamente las cosas se hacen más y más notorias y finalmente podremos ver algo donde antes no veíamos nada. Si la intensidad de la luz es muy baja, las cosas que vemos son *incoloras*. Es sabido que esta visión debida a la adaptación a la oscuridad se debe, por mucho, enteramente a los bastoncitos, mientras que la visión en la luz brillante se debe a los conos. Como resultado, existen muchos fenómenos que podemos apreciar fácilmente a causa de esta transferencia de funciones de los conos y bastoncitos en conjunto a los bastoncitos únicamente.

Hay muchas situaciones en las cuales, si la intensidad de la luz fuera más fuerte, podríamos ver color y encontraríamos estas cosas bien bellas. Un ejemplo es que a través del telescopio vemos casi siempre imágenes de nebulosas tenues en "blanco y negro", pero W. C. Miller, de los observatorios de Mt. Wilson y de Palomar, tuvo la paciencia de hacer fotografías *en color* de algunos de estos objetos. En realidad, nadie ha visto nunca estos colores con el ojo, pero no se trata de colores artificiales, se trata solamente de que la intensidad de la luz no es suficientemente fuerte para que los conos de nuestro ojo puedan verlos. Entre los más espectaculares de estos objetos están la nebulosa de anillo y la nebulosa del Cangrejo. La primera muestra una hermosa parte interna azul, con un halo rojo brillante exterior, y la última muestra una bruma general azulada, penetrada por filamentos rojo-anaranjados brillantes.

A la luz brillante, aparentemente, los bastoncitos son de sensibilidad muy baja, pero en la oscuridad, a medida que pasa el tiempo adquieren su habilidad para ver la luz. Las variaciones en intensidad de la luz para las cuales uno puede adaptarse son mayores que un millón a uno. La naturaleza no hace todo esto con sólo una clase de célula, sino que traspasa su tarea desde las células que ven la luz brillante, las células que ven los colores, los conos, a las células adaptadas a las intensidades bajas, a la oscuridad, los bastoncitos. Entre las consecuencias interesantes de este desplazamiento están, primero, que no existe color, y segundo, que hay una diferencia en la brillantez relativa de objetos coloreados de manera diferente. Resulta entonces que los bastoncitos ven mejor hacia el azul que los conos, y los conos pueden ver, por ejemplo, luz roja profunda, mientras que los bastoncitos la encuentran absolutamente imposible de ver. Así, pues, la luz roja es negra en lo que concierne a los bastoncitos. Por lo tanto, dos trozos de papel coloreado, digamos azul y rojo, donde el rojo podría ser aún más brillante que el azul bajo una buena luz, aparecerán, en la oscuridad, completamente invertidos. Se trata de

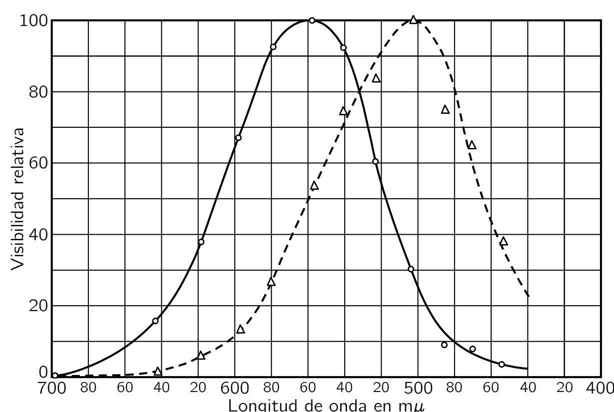


Figura 35-3. La sensibilidad espectral del ojo. Curva discontinua, bastoncitos; curva continua, conos.

un efecto bastante sorprendente. Si estamos en la oscuridad y podemos encontrar una revista o algo que tenga colores y, antes de que sepamos con seguridad cuáles son, juzgamos las áreas claras y oscuras, y si luego llevamos la revista a la luz, podríamos ver este corrimiento bien notable entre el que era el color más brillante y el que no lo era. El fenómeno se llama *efecto Purkinje*.

En la figura 35-3, la curva discontinua representa la sensibilidad del ojo en la oscuridad, es decir, utilizando los bastoncitos, mientras que la curva continua representa la sensibilidad a la luz. Vemos que la sensibilidad máxima de los bastoncitos está en la región verde y la de los conos está más en la región amarilla. Si se tiene una página de color rojo (el rojo está aproximadamente a 650 mμ) la podemos ver si está brillantemente iluminada, pero en la oscuridad es casi invisible.

Otro efecto del hecho de que los bastoncitos entran en juego en la oscuridad y de que no existen bastoncitos en la fóvea, es que en el momento que miramos directamente algo en la oscuridad, nuestra visión no es tan aguda como cuando miramos hacia un lado. Una estrella tenue, o una nebulosa, se puede ver a veces mejor mirando un poco hacia el lado en vez de directamente hacia ella, porque no tenemos bastoncitos sensibles en el medio de la fóvea.

Otro efecto interesante del hecho que el número de conos decrece a medida que vamos más hacia el lado del campo de visión, es que aun en una luz brillante, el color desaparece cuando el objeto se aleja hacia un lado. La manera de comprobar esto es mirar en alguna dirección fija particular, hacer que un amigo entre desde un lado con tarjetas coloreadas, y tratar de definir de qué color son antes de que estén exactamente frente a usted. Se encuentra que uno puede ver que las cartas están allí mucho antes de que se pueda determinar el color. Al hacer esto, es aconsejable entrar desde el lado opuesto del punto ciego, porque de otra manera resulta bastante confuso ver casi el color, entonces no ver nada y después ver el color otra vez.

Otro fenómeno interesante es que la periferia de la retina es muy sensible al movimiento. A pesar de que no podemos ver muy bien con el ángulo del ojo, si un pequeño bicho se mueve y nosotros no esperábamos que algo se moviera allí, somos inmediatamente sensibles a ello. Dentro de nosotros todo "se conecta" para mirar algo que se agita a un lado del campo.

35-3 Medición de la sensación de color

Ahora pasamos de la visión mediante los conos, a la visión de lo brillante, y llegamos al asunto que es más característico de la visión mediante los conos, y éste es el color. Como sabemos, la luz blanca se puede descomponer por medio de un prisma en todo un espectro de longitudes de onda que nos parecen tener diferentes colores; esto es lo que son los colores, por supuesto: apariencias. Cualquier fuente luminosa se puede analizar por medio de una red de difracción o un prisma y se puede determinar la distribución espectral, es decir, la "cantidad" de cada

longitud de onda. Una cierta luz puede tener mucho azul, bastante rojo y muy poco amarillo, o cualquier otra combinación. Eso es todo muy preciso en el sentido de la física, pero el problema es ¿de qué *color* aparecerá la luz? Es evidente que los diferentes colores dependen de alguna manera de la distribución espectral de la luz, pero el problema es encontrar cuáles características de la distribución espectral producen las diversas sensaciones. Por ejemplo, ¿qué tenemos que hacer para obtener un color verde? Todos sabemos que podemos tomar simplemente una parte del espectro que sea verde. Pero ¿es ésta la *única* manera de obtener verde, o anaranjado, o cualquier otro color?

¿Hay más de una distribución espectral que produzca el mismo efecto visual aparente? La respuesta es definitivamente, *sí*. Existe un número muy limitado de efectos visuales, de hecho, una variedad tridimensional de ellos, como veremos en breve, pero existe un número infinito de curvas diferentes que podemos trazar para la luz que proviene de diferentes fuentes. Ahora bien, el asunto que debemos discutir es: ¿en qué condiciones aparecen diferentes distribuciones de luz como exactamente del mismo color para el ojo?

La técnica psicofísica más poderosa para juzgar colores es usar el ojo como *instrumento de cero*. Esto es, no tratamos de definir qué constituye una sensación verde, o de medir en qué circunstancias obtenemos una sensación verde, porque resulta que esto es extremadamente complicado. En su lugar estudiamos las condiciones en las cuales dos estímulos son *indistinguibles*. Entonces no tenemos que decidir si dos personas ven la misma sensación en circunstancias diferentes, sino únicamente que si para una persona dos sensaciones son las mismas, son también las mismas para otra. No tenemos que decidir, cuando uno ve algo verde, si lo que siente en su interior es lo mismo que siente en su interior alguien distinto, cuando ve algo verde; no sabemos nada acerca de esto.

Para ilustrar las posibilidades podemos usar una serie de cuatro lámparas proyectoras que tienen filtros y cuyas intensidades se pueden ajustar en forma continua sobre un amplio intervalo; una tiene un filtro rojo y produce una mancha de luz roja sobre la pantalla, la siguiente tiene un filtro verde y produce una mancha verde, la tercera tiene un filtro azul y la cuarta es un círculo blanco con una mancha negra en su centro. Ahora bien, si encendemos alguna luz roja y a continuación ponemos algo de verde, vemos que en la superficie de traslape se produce una sensación que no es lo que llamamos verde rojizo, sino que un nuevo color, amarillo en este caso particular. Cambiando las proporciones de rojo y de verde podemos pasar por varios matices de anaranjado, etc. Si hemos ajustado un cierto amarillo, podemos obtener también el mismo amarillo no mezclando estos dos colores sino mezclando algunos otros, quizás un filtro amarillo con luz blanca, o algo parecido, para obtener la misma sensación. En otras palabras, es posible formar varios colores de más de una manera, mezclando las luces de diversos filtros.

Lo que hemos descubierto recién se puede expresar analíticamente como sigue. Un amarillo particular, por ejemplo, se puede representar por un cierto símbolo Y, que es la "suma" de ciertas porciones de luz roja filtrada (R) y luz verde filtrada (V). Usando dos números, digamos r y v para describir la intensidad de R y V, podemos escribir una fórmula para este amarillo:

$$Y = rR + vV. \tag{35.1}$$

El problema es: ¿podemos formar *todos* los diferentes colores sumando dos o tres luces de diferentes colores fijos? Veamos qué se puede hacer en relación con esto. Ciertamente no podemos obtener todos los diferentes colores mezclando sólo rojo y verde, porque, por ejemplo, el azul nunca aparece en tal mezcla. Sin embargo, introduciendo algo de azul, la región central, donde las tres regiones traslapan, se puede hacer aparecer de un blanco bastante hermoso. Mezclando los diversos colores y observando esta región central, encontramos que podemos obtener una gama considerable de colores en esta región cambiando las proporciones, y así no es imposible que *todos* los colores se puedan formar mezclando estas tres luces coloreadas. Discutiremos hasta donde esto es verdadero; es en efecto esencialmente correcto y veremos en breve cómo definir mejor ese enunciado.

Para ilustrar nuestro propósito, movamos las manchas sobre la pantalla de manera que todas caigan una sobre la otra y tratemos entonces de aparear con un color particular que aparece en el anillo producido por la cuarta lámpara. Lo que antes pensábamos que era “blanco” al salir de la cuarta lámpara, parece ahora amarillento. Podemos tratar de aparear éste, ajustando el rojo y el verde y el azul lo mejor que podamos en una especie de aproximaciones sucesivas y encontrar que podemos aproximarnos bastante a este matiz particular de color “crema”. Así, pues, no es difícil convencerse de que podemos formar todos los colores. Trataremos de formar el amarillo en breve, pero antes de hacer esto hay un color que podría ser difícil formar. Las personas que dan clases sobre los colores, forman todos los colores “brillantes”, pero nunca forman el *castaño*, y es difícil acordarse de haber visto alguna vez luz castaña. De hecho, este color no se usa nunca para algún efecto escénico, uno no ve nunca un reflector con luz castaña, así que pensamos que pudiera ser imposible formar el castaño. Para averiguar si es posible formar castaño, señalemos que luz castaña es simplemente algo que no estamos acostumbrados a ver sin su fondo. De hecho, podemos formarla mezclando algo de rojo y amarillo. Para demostrar que estamos observando luz castaña, aumentamos simplemente la luminosidad del fondo anular contra el cual vemos exactamente la misma luz ¡y vemos que ésta es, en efecto, lo que llamamos castaño! El castaño es siempre un color oscuro al lado de un fondo más luminoso. Podemos cambiar fácilmente la característica del castaño. Por ejemplo, si eliminamos algo de verde, obtenemos un castaño rojizo, aparentemente un castaño rojizo chocolate, y si agregamos más verde, en proporción, obtenemos aquel horrible color de que están hechos todos los uniformes del ejército, pero la luz de este color no es tan horrible en sí misma, es de un verde amarillento, pero vista contra un fondo luminoso.

Ahora colocamos un filtro amarillo frente a la cuarta luz y tratamos de aparearla. (La intensidad debe estar, por supuesto, dentro de las posibilidades de las diversas lámparas; no podemos aparear algo que sea demasiado luminoso, porque no tenemos suficiente potencia en la lámpara.) Pero *podemos* aparear el amarillo; usamos una mezcla de verde y de rojo y ponemos un toque de azul para hacerlo aún más perfecto. Estamos quizá listos para creer que, en buenas condiciones, podemos hacer un perfecto apareo de cualquier color dado.

Discutamos ahora las leyes de la mezcla de colores. En primer lugar, encontramos que distribuciones espectrales diferentes pueden producir el mismo color; a continuación, vimos que “cualquier” color puede formarse mezclando tres colores especiales, rojo, azul y verde. El aspecto más interesante de la mezcla de colores es este: si tenemos una cierta luz, que podríamos llamar X, y si ella aparece al ojo indistinguible de Y (puede ser una distribución espectral diferente, pero *aparece* como indistinguible), llamamos estos colores “iguales”, en el sentido de que los ojos los ven como iguales, y escribimos

$$X = Y. \tag{35.2}$$

Aquí está una de las grandes leyes del color: si dos distribuciones espectrales son indistinguibles, y si *agregamos a cada una* una cierta luz, digamos Z (si escribimos $X + Z$, esto significa que iluminamos ambas lámparas sobre la misma mancha) y tomamos entonces Y y agregamos la misma cantidad de la misma luz, Z, *las nuevas mezclas son también indistinguibles*:

$$X + Z = Y + Z. \tag{35.3}$$

Hemos apareado recién nuestro amarillo; si ahora iluminamos todo con luz rosada, todavía habrá apareo. Así, agregando cualquier otra luz a las luces apareadas, quedan apareadas. En otras palabras, podemos resumir todos estos fenómenos de color, diciendo que una vez que se tenga apareo entre dos luces coloreadas, vista una próxima a la otra en las mismas circunstancias, entonces este apareo permanecerá, y una luz puede ser sustituida por la otra luz en cualquier otra situación de mezcla de colores. En efecto, resulta, y esto es muy importante e interesante, que este apareo de color de las luces no depende de las características

del ojo en el momento de la observación: sabemos que si miramos por largo tiempo una superficie roja brillante, o una luz roja brillante, y miramos a continuación un papel blanco, éste se ve verdoso, y otros colores también se distorsionan, por haber estado mirando tanto tiempo el rojo brillante. Si tenemos ahora un apareo entre, digamos, dos amarillos ¡los miramos y los apareamos, enseguida miramos una superficie roja brillante por largo tiempo, y luego volvemos al amarillo, ya no se verá más amarillo; no sé de qué color se verá, pero no se verá amarillo. Sin embargo, *los amarillos todavía aparecerán apareados*, y así, a medida que el ojo se adapta a los diversos niveles de intensidad, el apareo de colores todavía funciona con la excepción evidente de cuando vamos a una región donde la intensidad de la luz se hace tan baja que nos hemos corrido de los conos a los bastoncitos, entonces el apareo de colores no será más un apareo de colores, porque estamos usando un sistema diferente.

El segundo principio de la mezcla de colores es éste: *cualquier color puede formarse a partir de tres colores diferentes*, en nuestro caso luces roja, verde y azul. Mezclando convenientemente las tres, podemos hacer cualquier cosa, como demostramos con nuestros dos ejemplos. Además, estas leyes son muy interesantes desde el punto de vista matemático. Para los que están interesados en las matemáticas del asunto, resulta como sigue. Suponga que tomamos nuestros tres colores que son rojo, verde y azul, pero los indicamos por A, B y C, y los llamamos nuestros colores *primarios*. Entonces cualquier color puede formarse con ciertas cantidades de estos tres: digamos una cantidad a de color A, una cantidad b de color B y una cantidad c de color C forma X:

$$X = aA + bB + cC. \tag{35.4}$$

Suponga ahora que se forma otro color Y a partir de los mismos tres colores:

$$Y = a'A + b'B + c'C. \tag{35.5}$$

Resulta entonces que la mezcla de las dos luces (esto es una de las consecuencias de las leyes que ya hemos mencionado) se obtiene haciendo la suma de las componentes de X y Y:

$$Z = X + Y = (a + a')A + (b + b')B + (c + c')C. \tag{35.6}$$

Es precisamente como la matemática de la suma de vectores, donde (a, b, c) son las componentes de un vector y (a', b', c') son las de otro vector, y la nueva luz Z es entonces la "suma" de los vectores. Este tema ha interesado siempre a los físicos y matemáticos. De hecho, Schrödinger escribió un maravilloso trabajo sobre la visión de los colores, en el cual desarrolló esta teoría del análisis vectorial aplicado a la mezcla de colores.†

Un problema se presenta ahora: ¿cuáles son los colores primarios correctos a usar? No hay tal cosa como "los" colores primarios correctos para la mezcla de luces. Podría tener, por razones prácticas, tres pinturas que son más útiles que otras para obtener una mayor variedad de pigmentos mezclados, pero no estamos discutiendo esta materia ahora. *Tres luces diferentemente coloreadas cualesquiera** pueden mezclarse siempre en la proporción correcta para producir *cualquier color*. ¿Podemos demostrar este hecho fantástico? En lugar de usar rojo, verde y azul, usemos rojo, azul y amarillo en nuestro proyector. ¿Podemos usar rojo, azul y amarillo para formar, digamos, verde?

Al mezclar estos tres colores en varias proporciones, obtenemos toda una gama de diferentes colores, abarcando todo un espectro. Pero de hecho, después de muchos tanteos, no encontramos nunca nada similar al verde. El problema es: ¿*podemos* formar el verde? La respuesta es sí. ¿Cómo? *Proyectando algo de rojo sobre el verde*, ¡entonces podemos hacer un emparejamiento con una cierta

† *Über das Verhältnis der Vierfarben- zur Dreifarbentheorie*, Sitzungsberichte der Akademie der Wissenschaften in Wien. Mathematisch-naturwissenschaftliche Klasse, Abteilung 2a, 134, (1925), 471-490.

* Excepto, por supuesto, si alguno de los tres se puede aparear mezclando los otros dos.

mezcla de amarillo y azul! Así los tenemos apareados, excepto que tuvimos que hacer trampa, introduciendo el rojo al otro lado. Sin embargo, ya que tenemos cierto refinamiento matemático, podemos darnos cuenta de que lo que realmente demostramos no fue que X puede formarse siempre, digamos a partir de rojo, azul y amarillo, sino que al introducir el rojo al otro lado, encontramos que rojo más X podía formarse a partir de azul y amarillo. Poniéndolo en el otro miembro de la ecuación, podemos interpretar eso como una *cantidad negativa*; así que, si admitimos que los coeficientes en ecuaciones como la (35.4) pueden ser tanto positivos como negativos, y si interpretamos las cantidades negativas como que tenemos que *sumar* ésas en el otro *miembro*, entonces cualquier color se puede aparear mediante tres cualesquiera, y no hay tal cosa como "los" primarios fundamentales.

Podemos preguntarnos si existen tres colores que sólo aparezcan con cantidades positivas para todas las mezclas. La respuesta es no. Todo conjunto de tres primarios requiere cantidades negativas para algunos colores, y por eso no existe una manera única de definir un primario. En los libros elementales se dice que son el rojo, el verde y el azul, pero esto es sólo porque con ésos se dispone de una *gama más amplia* de colores sin signos menos para algunas combinaciones.

35-4 El diagrama cromático

Discutamos ahora la combinación de colores a un nivel matemático, como una proposición geométrica. Si se representa cualquier color por la ecuación (35.4), podemos graficarlo como un vector en el espacio, trazando sobre los tres ejes las cantidades a, b y c y entonces cierto color será un punto. Si otro color es a', b', c', este color estará ubicado en alguna otra parte. La suma de ambos, como sabemos, es el color que resulta de sumarlos como vectores. Podemos simplificar este diagrama y representar todo en un plano, basándonos en la siguiente observación: si tuviéramos una cierta luz de color, y duplicáramos simplemente a, b y c, esto es, si los hacemos todos más intensos en la misma proporción, será el mismo color, pero más brillante. Por lo tanto, si convenimos en reducir todo a la *misma intensidad luminosa*, podemos proyectar todo sobre un plano, y esto es lo que se ha hecho en la figura 35-4. Resulta que cualquier color que se obtenga, mezclando otros dos en alguna proporción, estará en alguna parte sobre la línea trazada entre los dos puntos. Por ejemplo, una mezcla por mitades aparecería a la mitad entre ellos, y 1/4 de uno y 3/4 del otro aparecería a 1/4 de camino de un punto al otro, y así sucesivamente. Si usamos un azul, un verde y un rojo como primarios, vemos que todos los colores que podemos formar con coeficientes positivos están en el interior del triángulo de trazos, que contiene casi todos los colores que podamos ver alguna vez, porque todos los colores que podamos ver alguna vez están encerrados en el área de forma extraña, limitada por la curva. ¿De dónde proviene esta área? De una vez que alguien alza un apareo muy cuidadoso de todos los colores que podemos ver en relación con otros tres especiales. Pero no tenemos que comprobar *todos* los colores que podemos ver, tenemos que comprobar sólo los colores espectrales puros, las líneas del espectro. Una luz cualquiera se puede considerar como suma de diversas cantidades positivas de distintos colores espectrales puros —puros desde el punto de vista físico—. Una luz dada tendrá una cierta cantidad de rojo, amarillo, azul, etc. —colores espectrales—. Así, si sabemos cuánto de cada uno de nuestros tres primarios escogidos se necesita para formar cada uno de estos componentes puros, podemos calcular cuánto de cada uno se necesita para formar nuestro color dado. Así, si averiguamos cuáles son los *coeficientes de color* de todos los colores espectrales para tres colores primarios dados cualesquiera, podemos elaborar la tabla completa de mezcla de colores.

Un ejemplo de tales resultados experimentales al mezclar tres luces se da en la figura 35-5. Esta figura muestra la cantidad de cada uno de los tres diferentes primarios particulares, rojo, verde y azul, que se necesita para formar cada uno de los colores espectrales. El rojo está en el extremo izquierdo del espectro, el

Figura 35-4. El diagrama cromático normal.

Figura 35-5. Los coeficientes de color de los colores espectrales puros en términos de un cierto conjunto de colores primarios de referencia.

amarillo es el siguiente y así sucesivamente, hasta llegar al azul. Nótese que en algunos puntos se necesitan signos menos. Es a partir de tales datos que es posible ubicar la posición de cada uno de los colores en un diagrama, donde las coordenadas x y y están relacionadas con las cantidades de los diferentes primarios que se utilizan. Ésa es la manera como se ha encontrado la línea curva de contorno. Es el lugar geométrico de los colores espectrales puros. Ahora bien, cualquier otro color se puede formar sumando líneas espectrales, por supuesto, y así encontramos que todo lo que se pueda producir, uniendo una parte de esta curva con la otra, es un color que se encuentra en la naturaleza. La línea recta une el extremo violeta del espectro con el extremo rojo. Es el lugar geométrico de los púrpuras. En el interior del contorno están los colores que se pueden formar mediante luces, y en el exterior están los colores que no se pueden formar mediante luces, y nadie los ha visto nunca (¡excepto, posiblemente, como imágenes residuales!).

35-5 El mecanismo de la visión de los colores

El siguiente aspecto del problema es ahora la pregunta: *¿por qué* los colores se comportan de esta manera? La teoría más simple, propuesta por Young y Helmholtz, supone que en el ojo hay tres pigmentos diferentes que reciben la luz y tienen espectros de absorción distintos, de manera que un pigmento absorbe fuertemente, en el rojo, otro absorbe fuertemente en el azul, otro absorbe en el verde. Entonces, cuando hacemos incidir luz sobre ellos, vamos a obtener cantidades diferentes de absorción en las tres regiones y estas tres partes de información son manejadas de alguna manera en el cerebro o en el ojo, o en alguna parte, para decidir cuál es el color. Es fácil demostrar que todas las reglas de mezcla de colores serían una consecuencia de esta proposición. Ha habido grandes debates acerca del tema porque el problema siguiente, naturalmente, es encontrar las características de absorción de cada uno de los tres pigmentos. Resulta, desgraciadamente, que debido a que podemos transformar las coordenadas del color como queramos, sólo podemos encontrar todo tipo de combinaciones lineales de las curvas de absorción mediante experimentos de mezcla de colores, pero no las curvas para los pigmentos individuales. Se ha tratado de varias maneras de obtener una curva específica que describa alguna propiedad física particular del ojo. Una de estas curvas se llama *curva de luminosidad*, mostrada en la figura 35-3. En esta figura hay dos curvas, una para los ojos en la oscuridad, la otra para ojos a la luz, la última es la curva de luminosidad del Cono. Esto se mide encontrando cuál es la menor cantidad de luz de color que necesitamos para apenas verla. Esto mide la sensibilidad del ojo en las diferentes regiones espectrales. Hay otra manera muy interesante de medir esto. Si tomamos dos colores y los hacemos aparecer en una superficie, haciendo aparecer alternativamente uno tras otro, vemos un centelleo si la frecuencia es muy baja. Sin embargo, a medida que la frecuencia aumenta, el centelleo va a desaparecer por último a una cierta frecuencia que depende de la luminosidad de la luz, digamos a 16 repeticiones por segundo. Ahora bien, si ajustamos el brillo o la intensidad de un color respecto al otro, aparece una intensidad para la cual el centelleo de 16 ciclos desaparece. Para obtener centelleo con la luminosidad así ajustada, tenemos que ir a una frecuencia mucho más baja para ver un centelleo de color. Así, obtenemos lo que llamamos un centelleo de la luminosidad a una frecuencia más alta y, a una frecuencia más baja, un centelleo del color. Es posible ajustar dos colores a "luminosidad igual" por medio de esta técnica del centelleo. Los resultados son casi, pero no exactamente, los mismos que los obtenidos midiendo el umbral de sensibilidad del ojo para ver luz débil por los conos. La mayoría de los investigadores usan el sistema centelleo como una definición de la curva de luminosidad.

Ahora bien, si hay tres pigmentos sensibles al color en el ojo, el problema es determinar la forma del espectro de absorción de cada uno. ¿Cómo? Sabemos que existen personas que sufren de daltonismo —8 % de la población masculina y 0.5 % de la población femenina—. La mayoría de las personas que sufren de daltonismo o de alguna anormalidad en la visión del color tienen un grado

diferente de sensibilidad que los otros a una variación del color, pero de todas maneras necesitan tres colores para aparear. Sin embargo, hay algunos a los que se les llama *dicrómatas*, para quienes cualquier color se puede aparear usando solamente *dos* colores primarios. La sugerencia evidente es, entonces, decir que están perdiendo uno de los tres pigmentos. Si podemos encontrar tres tipos de dicrómatas daltónicos que tienen diferentes reglas para mezclar colores, un tipo debe estar perdiendo la pigmentación *roja*, la *verde* y otro la *azul*. ¡Midiendo todos estos tipos podemos determinar las tres curvas! Resulta que *hay* tres tipos de dicromatía daltónica; existen dos tipos comunes y un tercer tipo muy escaso, y a partir de estos tres ha sido posible deducir los espectros de absorción de los pigmentos.

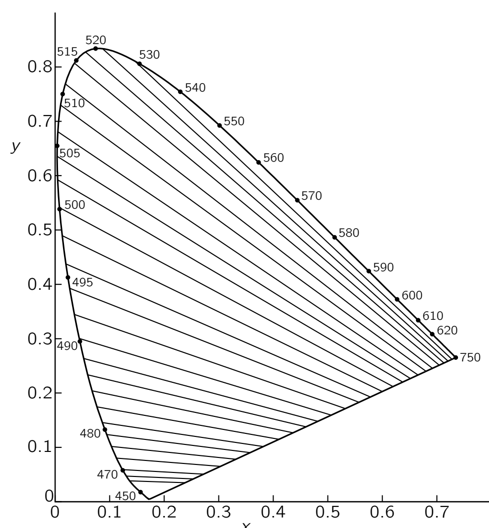


Figura 35-6. El lugar geométrico de los colores confundidos por los daltónicos deuteranopes.

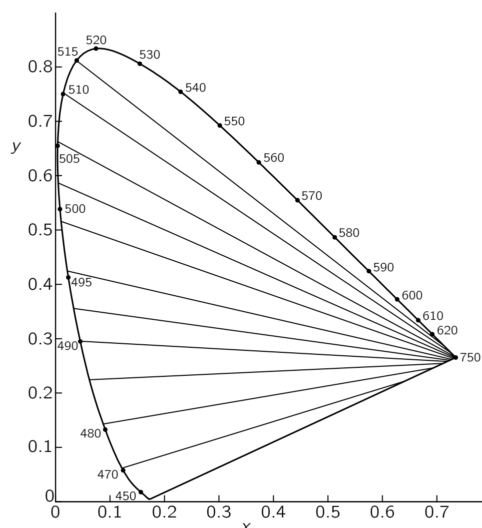


Figura 35-7. El lugar geométrico de los colores confundidos por los daltónicos protanopes.

La figura 35-6 muestra la mezcla de color de un tipo particular de persona daltónica llamado deuteranope. Para él, el lugar geométrico de colores constantes no son puntos, sino ciertas líneas, a lo largo de cada una de las cuales el color le parece a él ser el mismo. Si la teoría de que él está perdiendo una de las tres partes de información es correcta; todas estas líneas se deberían cortar en un punto. Si medimos cuidadosamente en el gráfico, se cortan perfectamente. Es evidente, entonces, que esto ha sido hecho por un matemático ¡y no representa datos reales! Si revisamos los últimos trabajos con datos verdaderos, resulta que en el gráfico de la figura 35-6, el punto de enfoque de todas las líneas no está exactamente en el lugar correcto. Usando las líneas de la figura de más arriba, no podemos encontrar espectros razonables, necesitamos absorciones negativas y positivas en diferentes regiones. Pero usando los nuevos datos de Yustova, resulta que cada una de las curvas de absorción es positiva en todas partes.

La figura 35-7 muestra un tipo diferente de daltonismo, el de protanope, quien tiene un enfoque cerca del final rojo de la curva de contorno. Yustova obtiene aproximadamente la misma ubicación en este caso. Usando los tres tipos diferentes de daltonismo se han determinado finalmente las tres curvas de respuesta a los pigmentos, y se muestran en la figura 35-8. ¿En forma terminante? Quizás. *Existe* una duda acerca de si la idea de los tres pigmentos es correcta, de si el daltonismo resulta por la pérdida de uno de los pigmentos, y aun si los datos de mezcla de color en el daltonismo son correctos. Diferentes investigadores obtienen diferentes resultados. Este campo se encuentra grandemente en desarrollo todavía.

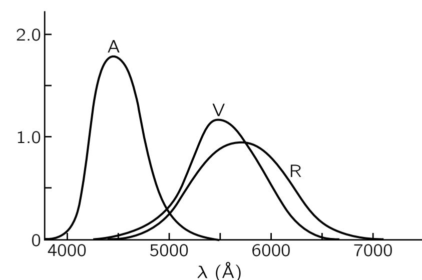


Figura 35-8. Las curvas de sensibilidad espectral de un receptor normal tricromático.

35-6 Fisicoquímica de la visión de los colores

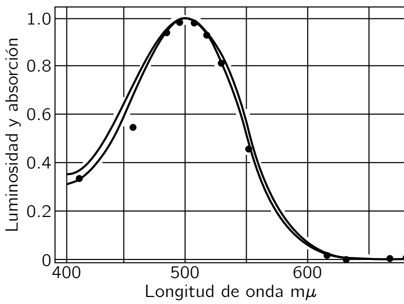


Figura 35-9. La curva de sensibilidad de un ojo adaptado a la oscuridad, comparada con la curva de absorción de la púrpura visual.

Ahora bien, ¿qué tal si verificamos estas curvas con los pigmentos reales en el ojo? Los pigmentos que se pueden obtener de una retina consisten principalmente de un pigmento llamado *púrpura visual*. Sus características más prominentes son, primero, que se encuentra en el ojo de casi todos los animales vertebrados, y segundo, que su curva de respuesta se ajusta maravillosamente con la sensibilidad del ojo, como se ve en la figura 35-9, en la cual están graficados en la misma escala la absorción del púrpura visual y la sensibilidad del ojo adaptado a la oscuridad. Éste es evidentemente el pigmento con el que vemos en la oscuridad: el púrpura visual es el pigmento para los bastoncitos y no tiene nada que ver con la visión del color. Este hecho fue descubierto en 1877. Aún hoy en día se puede decir que los pigmentos de color de los conos nunca han sido obtenidos en un tubo de ensayo. En 1958 se podía decir que los pigmentos de color nunca habían sido vistos. Pero desde entonces, dos de ellos han sido detectados por Rushton por medio de una técnica sencilla y hermosa.

La dificultad está, presumiblemente, en que, como el ojo es tan débilmente sensible a la luz brillante comparado con la luz de baja intensidad, necesita mucha púrpura visual para ver, pero no muchos de los pigmentos de color para ver los colores. La idea de Rushton es *dejar el pigmento en el ojo* y medirlo de todas maneras. Lo que hace es esto. Existe un instrumento llamado oftalmoscopio para enviar luz dentro del ojo a través de las lentes y luego enfocar la luz que regresa. Con él uno puede medir qué cantidad se refleja. Así, uno mide el coeficiente de reflexión de la luz que ha pasado *dos* veces a través del pigmento (reflejada por una capa posterior en el globo del ojo, y saliendo a través del pigmento del cono nuevamente). La naturaleza no está siempre tan maravillosamente diseñada. Los conos están diseñados en forma interesante para que la luz que llega a ellos rebote alrededor y busque su camino hacia abajo, hacia los pequeños puntos sensibles en el ápice. La luz baja derecho hacia el punto sensible, rebota en el fondo y regresa hacia afuera nuevamente, habiendo atravesado una cantidad considerable del pigmento de visión de color; también, examinando la fóvea, donde no existen bastoncitos, el púrpura visual no nos confunde. Pero el color de la retina se vio hace mucho tiempo: es algo como un rosado anaranjado; luego están todos los vasos sanguíneos y el color de la sustancia del fondo, y todo lo demás. ¿Cómo sabemos cuándo estamos mirando el pigmento? *Respuesta*: primero tomamos a una persona daltónica, quien tiene menos pigmentos y para quien, por lo tanto, es más fácil hacer el análisis. Segundo, los diversos pigmentos, como la púrpura visual, tienen un cambio de intensidad cuando son blanqueados por la luz; cuando los alumbramos cambian su concentración. Así, mientras miraba al espectro de absorción del ojo, Rushton introdujo *otro* rayo en todo el ojo, lo que cambia la concentración del pigmento, y midió el *cambio* en el espectro, y la diferencia, por supuesto, no tenía nada que ver con la cantidad de sangre o el color de las capas reflectoras, u otras cosas, sino solamente con el pigmento, y en esta forma Rushton obtuvo una curva para el pigmento del ojo protanope, la cual se da en la figura 35-10.

La segunda curva de la figura 35-10 es una curva obtenida con un ojo normal.

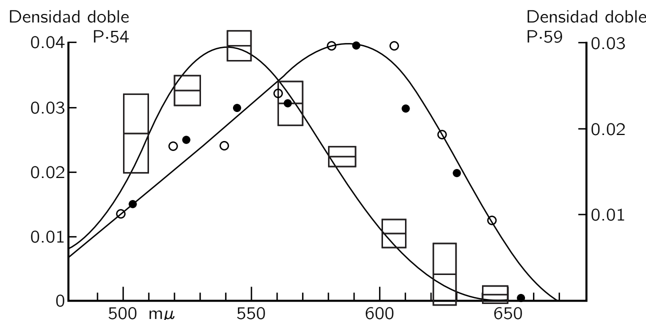


Figura 35-10. Espectro de absorción del pigmento del color de un daltónico protanope (cuadrados) y de un ojo normal (puntos).

Se obtuvo tomando un ojo normal y, habiendo ya determinado cuál era uno de los pigmentos iluminando el otro en el rojo, donde el primero es insensible. La luz roja no afecta el ojo protanope, pero sí el ojo normal, y así uno puede obtener la curva del pigmento que falta. La forma de una curva se ajusta maravillosamente a la curva del verde de Yustova, pero la curva del rojo está un poquito desplazada. Así que tal vez estamos en el buen camino. O tal vez no —los últimos trabajos con deuteranopes no demuestran que falte algún pigmento definido—.

El color no es un asunto de la física de la luz misma. El color es una *sensación*, y la sensación de los diferentes colores es diferente en distintas circunstancias. Por ejemplo, si tenemos una luz rosada, formada superponiendo haces cruzados de luz blanca y luz roja (todo lo que podemos hacer con el blanco y el rojo es rosado, evidentemente), podemos mostrar que la luz blanca puede aparecer azul. Si colocamos un objeto en los rayos da dos sombras —una iluminada por sólo la luz blanca y la otra por el rojo—. Para la mayoría de las personas la sombra "blanca" de un objeto se ve azul, pero si seguimos extendiendo esta sombra hasta que cubra toda la pantalla ¡vemos que de repente se ve blanca, no azul! Podemos obtener otros efectos de la misma naturaleza mezclando luces roja, amarilla y blanca. Las luces roja, amarilla y blanca pueden producir solamente amarillos anaranjados, etc. Así, pues, si mezclamos estas luces aproximadamente en forma igual, obtenemos solamente luz anaranjada. Sin embargo, al proyectar diferentes tipos de sombras en la luz, con distintos colores traslapados, uno obtiene una serie bastante grande de hermosos colores que no están en la luz misma (que es solamente anaranjada), sino en nuestras *sensaciones*. *Vemos* claramente muchos colores distintos que son totalmente diferentes a los colores "físicos" en el rayo. Es muy importante apreciar que una retina está ya "pensando" en la luz, está comparando lo que ve en una región con lo que ve en otra, aunque no en forma consciente. Lo que sabemos acerca de cómo lo hace es el tema del próximo capítulo.

BIBLIOGRAFÍA

Committee on Colorimetry, Optical Society of America, *The Science of Color*, Thomas Y. Crowell Company, Nueva York, 1953.

Hecht, S., S. Shlaer y M. H. Pirenne, "Energy, Quanta, and Vision", *Journal of General Physiology*, 1942, 25, 819-840.

Morgan, Clifford, y Eliot Stellar, *Physiological Psychology*, 2a. ed., McGraw-Hill Book Company, Inc., 1950.

Nuberg, N. D., y E. N. Yustova, "Researches on Dichromatic Vision and the Spectral Sensitivity of the Receptors of Trichromats", conferencia en el simposio núm. 8, Visual Problems of Colour, vol. II, National Physical Laboratory, Teddington, Inglaterra, septiembre de 1957. Publicado por Her Majesty's Stationery Office, Londres, 1958.

Rushton, W. A., "The Cone Pigments of the Human Fovea in Colour Blind and Normal", conferencia en el simposio núm. 8, Visual Problems of Colour, vol. I, National Physical Laboratory, Teddington, Inglaterra, septiembre de 1957. Publicado por Her Majesty's Stationery Office, Londres, 1958.

Woodworth, Robert S., *Experimental Psychology*, Henry Holt and Company, Nueva York, 1938. Edición revisada por Robert S. Woodworth y H. Schlosberg, 1954.

36

El mecanismo de la visión

36-1 La sensación del color

Al discutir el sentido de la vista, debemos darnos cuenta de que (¡fuera de una galería de arte moderno!) uno ve manchas de color o manchas de luz al azar. Cuando miramos un objeto vemos un *hombre* o una *cosa*; en otras palabras, el cerebro interpreta lo que vemos. Cómo lo hace, nadie lo sabe, y lo hace, por supuesto, a un nivel muy alto. Aunque evidentemente aprendemos a reconocer después de mucha experiencia cuál es el espacio de un hombre, hay una cierta cantidad de rasgos de la visión que son más elementales, pero que también implican combinar información de diferentes partes de lo que vemos. Para ayudarnos a entender cómo hacemos una interpretación de una imagen completa, vale la pena estudiar las primeras etapas de la unión de la información desde las diferentes células de la retina. En este capítulo nos vamos a dedicar principalmente a ese aspecto de la visión, aunque también vamos a mencionar algunos problemas laterales en el camino.

Un ejemplo del hecho de que tenemos acumulación de información muy elemental, desde diferentes partes del ojo al mismo tiempo, más allá de nuestro control voluntario o habilidad para aprender, fue esa sombra azul producida por la luz blanca, cuando tanto blanco como rojo estaban iluminando la misma pantalla. Este efecto por lo menos implica el conocimiento de que el fondo de la pantalla es rosado, aunque cuando estamos mirando a la sombra azul, es sólo luz "blanca" llegando a un punto particular en el ojo, en alguna parte, las informaciones han sido reunidas. Mientras más completo y familiar es el contexto, más correcciones hará el ojo para las partículas. De hecho, Land ha demostrado que si mezclamos ese azul aparente y el rojo en diferentes proporciones, mediante el uso de dos transparencias fotográficas que tienen diferente absorción frente al rojo y al blanco, se puede hacer que represente una escena real, con objetos reales, bastante fielmente. En este caso obtenemos también muchos colores aparentes intermedios, análogo a lo que obtendríamos al combinar rojo y azul verdoso, parece que fuera un conjunto casi completo de colores, pero si los miramos mucho, vemos que no son tan buenos. Aun así, es sorprendente todo lo que podemos obtener de sólo el rojo y el blanco. Mientras más se parece la escena a una situación real, ¡tanto más uno puede compensar el hecho de que toda la luz no es realmente nada más que rosado!

Otro ejemplo es la aparición de "colores" en un disco blanco y negro en rotación, cuyas áreas negras y blancas se muestran en la figura 36-1. Cuando el disco se rota, las variaciones de luz y sombra en cualquier radio son exactamente iguales; es sólo el fondo el que es diferente para las dos clases de "franjas". Sin embargo, uno de los "anillos" aparece coloreado con un color y el otro con otros.* Nadie entiende aún la razón de estos colores, pero está claro que la información es reunida a un nivel muy elemental, en el ojo mismo probablemente.

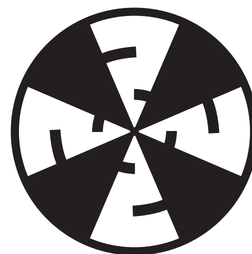


Figura 36-1. Cuando se hace rotar un disco como el de arriba, aparecen colores en sólo uno de los dos "anillos" más oscuros. Si se invierte el sentido de rotación, los colores aparecen en el otro anillo.

Casi todas las teorías modernas de la visión de los colores están de acuerdo en que los datos de mezcla de colores indican que hay sólo tres pigmentos en los conos del ojo y que es la absorción espectral en estos tres pigmentos lo que fundamentalmente produce el sentido del color. Pero la sensación total asociada con las características de absorción de los tres pigmentos actuando conjuntamente no es, necesariamente, la suma de las sensaciones individuales. Todos estamos de acuerdo en que el amarillo *no parece* ser verde rojizo; de hecho podría ser una

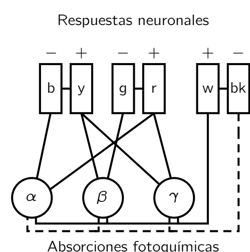
* Los colores dependen de la velocidad de rotación, de la intensidad de la iluminación y hasta cierto punto de quién los mira y qué tanto los observa.

tremenda sorpresa para mucha gente descubrir que la luz es, en realidad, una mezcla de colores, porque probablemente la sensación de luz se debe a algún otro proceso distinto al de una simple mezcla como un acorde en música, donde las tres notas están ahí al mismo tiempo y, si escuchamos con cuidado, podemos oírlas individualmente. No podemos mirar con cuidado y ver el rojo y el verde.

Las primeras teorías de la visión decían que hay tres pigmentos y tres tipos de conos, cada tipo conteniendo un pigmento, que un nervio va desde cada Cono al cerebro, de manera que las tres partes de la información son llevadas al cerebro, y entonces en el cerebro puede suceder cualquier cosa. Ésta, por supuesto, es una idea incompleta, no sirve de nada descubrir que el nervio óptico lleva la información al cerebro, porque aún no hemos empezado ni siquiera a resolver el problema. Debemos plantearnos preguntas más básicas: ¿Hace alguna diferencia *dónde* se junta la información? ¿Es importante que sea llevada directamente por el nervio óptico hasta el cerebro, o podría la retina hacer primero algún análisis? Hemos visto un cuadro de la retina como una cosa muy complicada, con muchas interconexiones (figura 35-2) y ella podría hacer algunos análisis.

De hecho, las personas que estudian anatomía y el desarrollo del ojo han demostrado que la retina es, en realidad, cerebro: en el desarrollo del embrión, un pedazo del cerebro sale hacia adelante y fibras largas crecen hacia atrás, conectando el ojo con el cerebro. La retina está organizada de la misma manera que está organizado el cerebro y, como lo ha dicho alguien en forma tan bella, "el cerebro ha desarrollado una manera de mirar hacia el mundo". El ojo es un pedazo de cerebro que está tocando la luz, por decirlo así, en el exterior. De manera que no sería raro que algún análisis del color ya se haya hecho en la retina.

Esto nos da una oportunidad muy interesante. Ninguno de los otros sentidos implica tal cantidad de cálculo, digamos, que uno pueda hacer medidas antes que la señal entre a un nervio. Los cálculos para todos los otros sentidos generalmente se realizan en el cerebro mismo, donde es muy difícil llegar a lugares específicos para hacer medidas, porque hay tantas interconexiones. Aquí, con el sentido de la vista, tenemos la luz, tres capas de células haciendo cálculos y los resultados de estos cálculos son transmitidos a través del nervio óptico. De manera que tenemos la primera oportunidad de observar fisiológicamente, quizá, cómo las primeras capas del cerebro trabajan en sus primeras etapas. Es así de doble interés, no solamente interesante para la visión, sino interesante para todo el problema de fisiología.



$$y - b = k_1(\beta+\gamma-2\alpha)$$
$$r - g = k_2(\alpha+\gamma-2\beta)$$
$$w - bk = k_3(\alpha+\gamma+\beta) - k_4(\alpha+\beta+\gamma)$$

Figura 36-2. Conexiones neuronales de acuerdo con una teoría "de los opuestos" de la visión del color.

El hecho de que haya tres pigmentos no significa que deba haber tres tipos de sensaciones. Otra de las teorías de la visión de los colores dice que hay esquemas de colores que realmente se oponen (figura 36-2). O sea, una de las fibras nerviosas lleva muchos impulsos si se está viendo amarillo y menos de lo usual para el azul. Otra fibra nerviosa lleva información verde y roja de la misma manera, y otra, blanca y negra. En otras palabras, en esta teoría alguien ha empezado ya a tratar de adivinar el circuito, el método de cálculo.

Los problemas que estamos tratando de resolver estimando estos primeros cálculos son problemas sobre los colores aparentes que se ven en un fondo rosado, qué pasa cuando el ojo se adapta a diferentes colores y también los así llamados fenómenos psicológicos. Los fenómenos psicológicos son del tipo, por ejemplo, que el blanco no "se siente" como rojo, amarillo y azul, y esta teoría fue propuesta porque los psicólogos dicen que hay *cuatro* colores puros aparentes: "Hay cuatro estímulos que tienen una gran capacidad para evocar psicológicamente los simples colores, azul, amarillo, verde y rojo, respectivamente. En contraste con el tierra de Siena, el magenta, el púrpura o la mayoría de los colores discriminables, estos colores no son mezclados en el sentido que ninguno comparte la naturaleza del otro, específicamente, el azul no es amarillento, ni rojizo o verdoso y así sucesivamente, son colores psicológicamente primarios". Éste es un hecho psicológico, así se llama. Para encontrar de qué evidencia se dedujo este hecho psicológico, debemos buscar en forma realmente afanosa en toda la literatura. En la literatura moderna, todo lo que encontramos sobre el tema son repeticiones de la misma afirmación o de una de un psicólogo alemán que usa como una de sus autoridades a Leonardo da

Vinci, quien, todos sabemos por supuesto, era un gran artista. Él dice: "Leonardo pensó que había cinco colores". Entonces, buscando aún más, encontramos en un libro aún más viejo la evidencia para el tema. El libro dice algo por este estilo: "El púrpura es azul rojizo, el naranja es amarillo rojizo, ¿pero puede el rojo verse como naranja purpúreo? ¿No son el rojo y el amarillo más unitarios que el púrpura y el naranja? Una persona promedio, a quien se le pide que diga cuáles colores son unitarios, nombra el rojo, el amarillo y el azul, estos tres, y algunos observadores, agregan un cuarto, el verde. Los psicólogos están acostumbrados a aceptar los cuatro colores salientes". Así que éste es el estado en el análisis psicológico de este tema, si todos dicen que hay tres y alguien dice que hay cuatro, y ellos quieren que sean cuatro, serán cuatro. Esto muestra la dificultad con las investigaciones psicológicas. Está claro que tenemos tales sensaciones, pero es muy difícil obtener mucha información acerca de ellas.

Por lo tanto, el otro camino a seguir es la dirección de la fisiología, para averiguar de manera experimental lo que realmente sucede en el cerebro, el ojo, la retina o donde sea, y tal vez descubrir que algunas combinaciones de impulsos de diversas células se mueven a lo largo de ciertas fibras nerviosas. A propósito, los pigmentos primarios no tienen que estar en células separadas, uno podría tener células en las cuales hay mezclas de los diversos pigmentos, células con los pigmentos rojos y verdes, células con los tres (la información de los tres es información blanca), etc. Hay muchas maneras de montar el sistema, y debemos encontrar cuál usó la naturaleza. Se podría esperar, en última instancia, que cuando entendamos las conexiones fisiológicas vamos a entender algo de algunos de estos aspectos de la psicología, así que buscaremos en esa dirección.

36-2 La fisiología del ojo

Empezamos hablando no sólo de la visión de los colores, sino también de la visión en general, sólo para recordarnos de las interconexiones que hay en la retina, mostradas en la figura 35-2. La retina es realmente como la superficie del cerebro. Aunque la fotografía real a través de un microscopio se ve un poco más complicada que este dibujo un tanto esquematizado, mediante un análisis cuidadoso, uno puede ver todas estas interconexiones. No hay duda que de una parte de la superficie de la retina está conectada a otras partes y que la información que sale de los largos axones, que producen el nervio óptico, son combinaciones de información de muchas células. Hay tres capas de células en la sucesión de funciones: las células de la retina, que son las que afecta la luz, una célula intermedia, que toma la información de una o unas pocas células de la retina y que la entrega a varias células en una tercera capa de células y la lleva al cerebro. Hay todo tipo de interconexión entre las células de las capas.

Ahora consideramos algunos aspectos de la estructura y comportamiento del ojo (véase figura 35-1). El enfoque de la luz es realizado principalmente por la córnea, debido a que tiene una superficie curva que "dobla" la luz. Por esto es que no podemos ver claramente bajo el agua, porque entonces no tenemos suficiente diferencia entre el índice de la córnea, que es 1.37, y el del agua, que es 1.33. Detrás de la córnea hay agua, prácticamente, con un índice 1.33 y más atrás hay una lente que tiene una estructura muy interesante, es una serie de capas como una cebolla con la diferencia de que es transparente y que tiene un índice 1.40 en el medio y 1.38 exteriormente. (Sería bueno si pudiéramos hacer un vidrio óptico en el cual pudiéramos ajustar el índice en todas partes, porque entonces no tendríamos que encorvarlo tanto como lo hacemos cuando tenemos un índice uniforme.) Además, la forma de la córnea no es la de una esfera. Una lente esférica tiene una cierta cantidad de aberración esférica. La córnea es "más plana" por afuera que una esfera, de manera tal ¡que la aberración esférica es menor para la córnea de lo que sería si pusiéramos una lente esférica ahí! La luz es enfocada por el sistema córnea-lente en la retina. Cuando miramos cosas que están más cerca y más lejos, la lente se aprieta y se suelta y cambia el foco para ajustarse a las diferentes distancias. Para ajustar la cantidad total de luz está el

iris, que es lo que llamamos el color del ojo, castaño o azul; depende de quien sea, a medida que la cantidad de luz aumenta y disminuye, el iris se cierra y se abre.

Observemos ahora el mecanismo nervioso para controlar la acomodación de la lente, el movimiento del ojo, los músculos que mueven los ojos en las órbitas y el iris, mostrado esquemáticamente en la figura 36-3. De toda la información que sale del nervio óptico A, la gran mayoría se distribuye en uno de dos manojos (de los cuales vamos a hablar después) y de ahí al cerebro. Pero hay unas cuantas fibras, de interés ahora para nosotros, que no van directamente a la corteza visual del cerebro donde "vemos" las imágenes, sino al cerebro medio H. Éstas son las fibras que miden la luz promedio y hacen ajustes para el iris, o si la imagen se ve borrosa tratan de corregir la lente, o si hay una imagen doble tratan de ajustar el ojo para visión binocular. En todo caso, atraviesan el cerebro medio y realimentan el ojo. En K están los músculos que gobiernan las acomodaciones de la lente y en L otro que va al iris. El iris tiene dos sistemas de músculos. Uno es un músculo circular L que cuando es excitado, tira y cierra el iris; actúa muy rápidamente y los nervios están conectados directamente desde el cerebro a través de cortos axones con el iris. Los músculos opuestos son músculos radiales, de manera que cuando oscurece y los músculos circulares se relajan, estos músculos radiales tiran hacia afuera. Aquí tenemos, como en muchos lugares del cuerpo, un par de músculos que trabajan en direcciones opuestas y en casi todos estos casos los sistemas nerviosos que controlan a ambos están ajustados muy delicadamente, de manera que cuando se envían señales para contraer uno, automáticamente se envían señales para dilatar el otro. El iris es una excepción peculiar, ya que los nervios que hacen contraerse el iris son los que hemos descrito, pero los nervios que hacen que el iris se *dilate*, nadie sabe exactamente de dónde salen: bajan por la espina dorsal detrás del pecho hasta las secciones torácicas, salen de la espina dorsal, suben a través de los ganglios del cuello y dando toda la vuelta retornan a la cabeza, de modo de accionar el otro extremo del iris. De hecho, la señal va por un sistema nervioso completamente diferente, no el sistema nervioso central en absoluto, sino el sistema nervioso simpático; así que ésta es una manera bastante extraña de hacer funcionar las cosas.

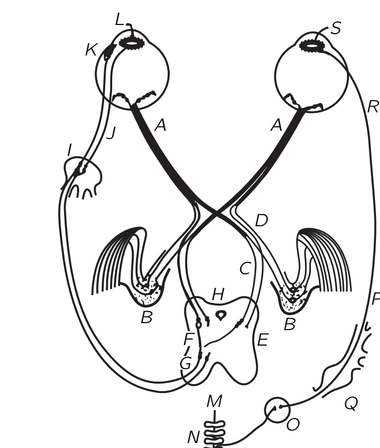


Figura 36-3. Las interconexiones neurales para el funcionamiento mecánico de los ojos.

Ya hemos señalado otra cosa extraña acerca del ojo, que las células sensibles a la luz están en el lado equivocado, de manera que la luz debe atravesar varias capas de otras células antes de llegar a los receptores —¡está construido al revés!—. De manera que algunos aspectos son maravillosos y otros aparentemente estúpidos.

La figura 36-4 muestra las conexiones del ojo con la parte del cerebro que está más directamente relacionada con el proceso visual. Las fibras del nervio óptico van a una cierta área justo más allá de D, llamada geniculado lateral, desde donde van a una sección del cerebro llamada corteza visual. Nótese que algunas de las fibras de cada ojo se mandan al otro lado del cerebro, de manera que el cuadro formado es incompleto. Los nervios ópticos del lado izquierdo del ojo derecho pasan a través del quiasma óptico B, mientras que los del lado izquierdo del ojo izquierdo dan la vuelta y van por este mismo camino. De manera que el lado izquierdo del cerebro recibe toda la información que viene del lado izquierdo del globo de cada ojo, es decir, en el lado derecho del campo visual, mientras que el lado derecho del cerebro ve el lado izquierdo del campo visual. Ésta es la manera como la información de los dos ojos se junta para decir a qué distancia están las cosas. Éste es el sistema de visión binocular.

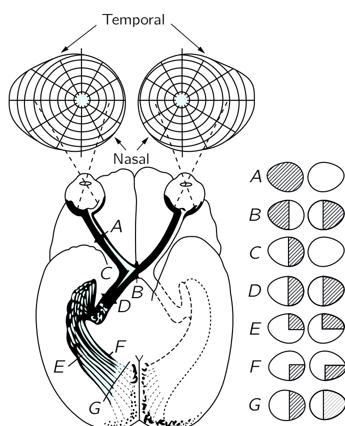


Figura 36-4. Las conexiones neurales de los ojos a la corteza visual.

Las conexiones entre la retina y la corteza visual son interesantes. Si un punto de la retina se saca o se destruye de alguna manera, entonces toda la fibra muere y podemos así encontrar dónde está conectada. Resulta que, esencialmente, las conexiones son una a una —por cada punto de la retina hay un punto de la corteza visual—, y puntos que están muy juntos en la retina están muy juntos en la corteza visual. De manera que la corteza visual todavía representa el arreglo espacial de los bastoncitos y conos, pero, por supuesto, muy distorsionado. Cosas que están en el centro del campo, que ocupan un lugar muy pequeño en la retina, están extendidas sobre muchas, muchas células en la corteza visual. Está claro que es útil tener cosas que estaban originalmente juntas, más juntas aún. El aspecto más notable del tema, sin embargo, es el siguiente. El lugar donde podría

pensarse que sería más importante tener cosas juntas sería exactamente en el medio del campo visual. Créanlo o no, la línea vertical en nuestro campo de visión cuando miramos algo es de tal naturaleza que la información de todos los puntos en el lado derecho de esa línea va al lado izquierdo del cerebro y la información de los puntos del lado izquierdo va al lado derecho del cerebro, y de la manera como esta región está hecha, hay un corte justo en el medio, de manera que las cosas que están muy juntas en el medio ¡están muy separadas en el cerebro! De alguna manera, la información tiene que ir desde un lado del cerebro al otro por algún otro conducto, lo que es bastante sorprendente.

El problema de cómo esta red llega a "armarse" es muy interesante. El problema de cuánto está ya armado y cuánto se aprende es muy antiguo. Se pensaba hace mucho que a lo mejor no necesitaba en absoluto estar armada cuidadosamente, que está interconectada sólo superficialmente y luego, mediante la experiencia, el niño aprende que cuando una cosa está "ahí arriba" produce alguna sensación en el cerebro. (Los doctores siempre nos dicen lo que "siente" un niño; pero, ¿cómo saben *ellos* lo que siente un niño a la edad de un año?) El niño, a la edad de un año, se supone que ve que un objeto está "ahí arriba", obtiene una cierta sensación y aprende cómo alcanzar hasta "ahí", porque cuando alcanza hasta "aquí", no resulta. Este enfoque probablemente no es correcto, porque ya vemos que en muchos casos están estas interconexiones especialmente detalladas. Muy esclarecedoras son algunas notables experiencias realizadas con salamandras. (A propósito, en la salamandra hay una conexión cruzada directa, sin el quiasma óptico, porque los ojos están a cada lado de la cabeza y no tienen región en común. Las salamandras no tienen visión binocular.) La experiencia es ésta: podemos cortar el nervio óptico de una salamandra y el nervio va a crecer de nuevo desde los ojos. Miles y miles de fibras celulares se regeneran así. Ahora bien, en el nervio óptico las fibras no están adyacentes —es como un gran cable telefónico hecho descuidadamente con todas las fibras torciéndose y dando vueltas—; pero cuando llegan al cerebro se ordenan de nuevo. Cuando cortamos el nervio óptico de la salamandra, la pregunta interesante es: ¿podrá ordenarse alguna vez? La respuesta es notable: sí. Si cortamos el nervio óptico de la salamandra y éste crece de nuevo, la salamandra tiene de nuevo buena agudeza visual. Sin embargo, si cortamos el nervio óptico y *damos vuelta al ojo de arriba para abajo*, y dejamos crecer aquél de nuevo, otra vez tendrá una agudeza visual correcta, pero con un terrible error: cuando la salamandra ve una mosca "ahí arriba" salta sobre ella "allá abajo" y nunca aprende. Por lo tanto, hay una manera misteriosa mediante la cual las miles y miles de fibras encuentran su lugar adecuado en el cerebro.

Este problema de cuánto está armado y cuánto no lo está, es un problema importante en la teoría del desarrollo de las criaturas. La respuesta no se conoce, pero se está estudiando intensivamente.

El mismo experimento, en el caso del pez dorado, demuestra que se produce un terrible nudo, como una gran cicatriz o complicación donde cortamos el nervio óptico; pero, a pesar de todo esto, las fibras crecen de nuevo a sus lugares correctos en el cerebro.

Para hacer esto, a medida que crecen en los viejos canales del nervio óptico, deben realizar una serie de decisiones respecto a la dirección en que deben crecer. ¿Cómo hacen esto? Parece que hay claves químicas a las cuales diferentes fibras responden en forma distinta. Piensen en el enorme número de fibras en crecimiento, cada una de las cuales es un individuo que difiere de alguna manera con sus vecinos; al responder a cualquiera que sean las claves químicas ¡responde de un modo único suficiente para encontrar su lugar apropiado para la conexión definitiva en el cerebro! Esto es un hecho interesante —fantástico—. Es uno de los grandes fenómenos de la biología descubierto recientemente y está conectado indudablemente con otros antiguos problemas no resueltos de crecimiento, organización y desarrollo de organismos y particularmente de embriones.

Otro fenómeno interesante tiene que ver con el movimiento del ojo. Hay que mover los ojos para hacer que las dos imágenes coincidan en diferentes circunstancias. Estos movimientos son de diferente tipo: uno es seguir algo, lo

que demanda que los dos ojos vayan en la misma dirección, hacia la derecha o la izquierda, y el otro es dirigirlos hacia el mismo lugar a distintas distancias, lo que precisa que se muevan en forma opuesta. Los nervios que van a los músculos del ojo ya están conectados para estos efectos. Hay un grupo de nervios que va a tirar los músculos del lado de dentro de un ojo y el de fuera del otro y relajar los músculos opuestos, de manera que los dos ojos se muevan juntos. Hay otro centro en el cual una excitación va a producir que los ojos se muevan uno hacia el otro desde una posición paralela. Cualquier ojo puede ser girado hacia la esquina si el otro ojo se mueve hacia la nariz, pero es imposible en forma consciente o inconsciente girar ambos ojos hacia *afuera*, al mismo tiempo, no porque no haya *músculos*, sino porque no hay manera de mandar una señal para girar ambos ojos hacia afuera, a menos que hayamos tenido un accidente o haya pasado algo, por ejemplo, que un nervio haya sido cortado. Aunque los músculos de un ojo pueden, ciertamente, dirigir un ojo para cualquier lado, ni siquiera un yogui es capaz de mover *ambos* ojos libremente hacia afuera bajo control voluntario, porque parece ser que no hay manera de hacerlo. Estamos ya armados hasta cierto punto. Éste es un punto importante, porque la mayoría de los anteriores libros de anatomía y de psicología, etc., no aprecian o no ponen énfasis en el hecho que ya estamos armados tan completamente —dicen que todo se aprende—.

36-3 Las células bastoncitos

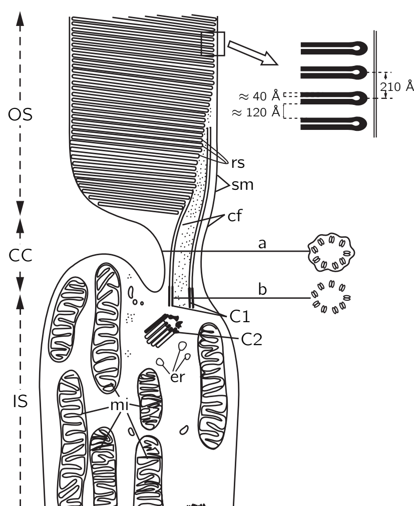


Figura 36-5. Microfotografía electrónica de una célula bastoncito.

Analicemos ahora con mayor detalle qué sucede en las células bastoncitos. La figura 36-5 muestra una microfotografía electrónica del medio de una célula bastoncitos (la célula bastoncito se extiende fuera del campo). Hay capa tras capa de estructuras planas, que se muestran ampliadas a la derecha, que contienen la sustancia rodopsina (púrpura visual), la tintura o pigmento, que produce los efectos de visión en los bastoncitos. La rodopsina, que es el pigmento, es una proteína grande que contiene un grupo especial llamado retineno, que se puede extraer de la proteína y que es, indudablemente, la causa principal de la absorción de la luz. No entendemos la razón de los planos, pero es muy posible que haya alguna razón para mantener paralelas todas las moléculas de rodopsina. La química del fenómeno ha sido estudiada extensamente, pero podría haber algo de física en él. Pudiera ser que todas las moléculas estén ordenadas en una especie de fila, de manera que cuando una se excita se genera un electrón, digamos, que puede recorrerlas todas hasta algún lugar al final para sacar la señal, o algo por el estilo. Este tema es muy importante y no ha sido estudiado. Es un campo en el cual tanto la bioquímica como la física del estado sólido, o algo parecido, se va a usar a la larga.

Este tipo de estructura en capas aparece en otras circunstancias donde la luz es importante, por ejemplo, en el cloroplasto de las plantas, donde la luz produce la fotosíntesis. Si los ampliamos, encontramos lo mismo, con casi los mismos tipos de capas, pero ahí tenemos clorofila, por supuesto, en vez de retineno. La forma química del retineno se muestra en la figura 36-6. Tiene una serie de enlaces dobles alternados en la cadena lateral que es característica de casi todas las sustancias orgánicas fuertemente absorbentes, como la clorofila, la sangre, etc. Es imposible que los seres humanos fabriquen esta sustancia en sus propias células —tenemos que comerla—. De manera que la comemos bajo la forma de una sustancia especial, que es exactamente igual al retineno, excepto que hay un hidrógeno amarrado en el extremo derecho; se llama vitamina A y si no comemos suficientemente de ella, no obtenemos el abastecimiento adecuado de retineno y el ojo adquiere lo que llamamos *ceguera nocturna*, porque entonces no hay suficiente pigmento en la rodopsina para ver de noche con los bastoncitos.

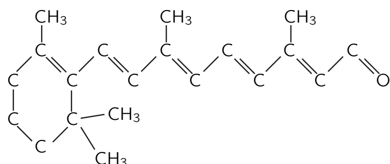


Figura 36-6. La estructura del retineno.

La razón por la cual tal serie de enlaces dobles absorbe luz muy fuertemente, también se conoce. Podemos dar una sugerencia: la serie alternante de enlaces dobles se llama enlace doble *conjugado*; un enlace doble significa que hay un electrón extra y éste se muda fácilmente hacia la derecha o la izquierda. Cuando la luz incide sobre esta molécula, el electrón de cada enlace doble se corre un paso. Todos los electrones en toda la cadena se corren como una fila de fichas de

dominó que se caen, y aunque cada uno se mueve sólo una pequeña distancia (es de esperar que, en un átomo simple, podamos mover el electrón sólo una pequeña distancia) ¡el efecto neto es el mismo que si el que está en un extremo se hubiera movido al otro extremo! Es lo mismo que si un electrón recorriera toda la distancia hacia adelante y hacia atrás y así, de esta manera, obtenemos una absorción mucho más fuerte bajo la influencia del campo eléctrico que si sólo pudiéramos mover el electrón una distancia que está asociada a un átomo. De manera que, como es fácil de mover los electrones hacia atrás y hacia adelante, el retineno absorbe la luz muy fuertemente; ése es el mecanismo de su parte físico-química.

36-4 El ojo compuesto (del insecto)

Volvamos a la biología. El ojo humano no es el único tipo de ojo. En los vertebrados, casi todos los ojos son esencialmente como los ojos humanos. Sin embargo, en los animales inferiores hay muchos otros tipos de ojos: manchas oculares, diferentes copas de ojos y otras cosas menos sensibles, que no tenemos tiempo de discutir. Pero hay otro ojo muy desarrollado en los invertebrados, el ojo *compuesto* del insecto. (La mayoría de los insectos que tienen grandes ojos compuestos, también tienen varios ojos adicionales más simples.) La abeja es un insecto cuya visión ha sido estudiada cuidadosamente. Es fácil estudiar las propiedades de la visión de las abejas, porque son atraídas por la miel y podemos hacer experimentos en los cuales identificamos la miel, poniéndola en papel azul o papel rojo y vemos a cuál vienen. Mediante este método se han descubierto algunas cosas muy interesantes respecto a la visión de la abeja.

En primer lugar, al tratar de medir con cuánta agudeza pueden las abejas ver la diferencia de color entre dos pedazos de papel "blanco", algunos investigadores encontraron que no era muy buena y otros encontraron que era fantásticamente buena. Aun cuando los dos pedazos de papel blanco eran casi exactamente iguales, las abejas podían distinguir la diferencia. Los investigadores usaron blanco de cinc para un pedazo de papel y blanco de plomo para otro y, aunque para nosotros se ven exactamente iguales, la abeja podía distinguirlos perfectamente, porque reflejan en cantidades diferentes el ultravioleta. De esta manera se descubrió que el ojo de la abeja es sensible a un intervalo mayor del espectro que los nuestros. Nuestro ojo trabaja entre 7 000 y 4 000 ángstroms, desde el rojo al violeta, pero la abeja alcanza a ver hasta 3 000 ángstroms ¡en el ultravioleta! Esto produce una serie de diferentes efectos interesantes. En primer lugar, las abejas pueden distinguir entre muchas flores que a nosotros nos parecen iguales. Por supuesto, debemos darnos cuenta de que los colores de las flores no están diseñados para *nuestros* ojos, sino para los de la abeja, son señales para atraer las abejas a una flor específica. Todos sabemos que hay muchas flores "blancas". Parece ser que el blanco no es muy interesante para las abejas porque resulta que todas las flores blancas tienen diferentes proporciones de reflexión en el *ultravioleta*; no reflejan el 100 % del ultravioleta, como lo haría un blanco verdadero. No toda la luz vuelve, falta el ultravioleta y ése es un color, como para nosotros cuando faltara el azul, parece amarillo. De manera que todas las flores tienen colores para las abejas. Sin embargo, también sabemos que las abejas no pueden ver el rojo. Por lo tanto, podríamos esperar que todas las flores rojas parecieran negras a la abeja. ¡De ningún modo! Un estudio cuidadoso de las flores rojas muestra, primero, que aun con nuestros ojos podemos ver que la mayoría de las flores rojas tienen un tinte azulado, principalmente porque están reflejando una cantidad adicional de azul, que es la parte que ven las abejas. Además, los experimentos demuestran también que las flores varían según su reflexión en el ultravioleta, en las diferentes secciones de los pétalos, etc. ¡De manera que si pudiéramos ver las flores como las ven las abejas, serían aún más bonitas y variadas!

Se ha demostrado, sin embargo, que hay unas cuantas flores rojas que *no* reflejan el azul o el ultravioleta y, por lo tanto, ¡*parecerían* negras a la abeja! Esto era de sumo interés para las personas que se preocupan de este tema, porque el negro no parece un color interesante, ya que es difícil distinguirlo de una vieja

sombra sucia. Resultó ser que estas flores *no* eran visitadas por las abejas, son las flores visitadas por los *colibríes* ¡y los colibríes *pueden* ver el rojo!

Otro aspecto interesante de la visión de la abeja es que las abejas pueden, aparentemente, indicar la dirección del Sol mirando un pedazo de cielo azul, sin ver el Sol mismo. Nosotros no podemos hacer esto fácilmente. Si miramos el cielo por la ventana hacia afuera y vemos que es azul, ¿en qué dirección está el Sol? La abeja puede decirlo, porque la abeja es muy sensible a la *polarización* de la luz y la luz dispersada del cielo está polarizada.* Todavía existe alguna discusión sobre cómo funciona esta sensibilidad. Aún no se sabe si es porque las reflexiones de la luz son diferentes en diversas circunstancias o porque el ojo de la abeja es sensible directamente.†

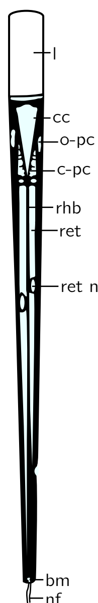


Figura 36-7. La estructura de un omatidio (una célula simple de un ojo compuesto).

También se dice que la abeja puede apreciar un centelleo de hasta 200 oscilaciones por segundo, mientras que nosotros no vemos más que hasta 20. Los movimientos de las abejas en las colmenas son muy rápidos; los pies se mueven y las alas vibran, pero es muy difícil para nosotros ver estos movimientos con nuestros ojos. Sin embargo, si pudiéramos ver más rápidamente veríamos el movimiento. Probablemente es muy importante para la abeja que su ojo tenga una respuesta tan rápida.

Discutamos ahora qué agudeza visual podríamos esperar de la abeja. El ojo de la abeja es compuesto y está hecho de un gran número de células especiales llamadas *omatidios*, que están arregladas en forma cónica en la superficie de una esfera (aproximadamente) en la parte de fuera de la cabeza de la abeja. La figura 36-7 muestra un dibujo de uno de estos omatidios. En la parte de arriba hay un área transparente, un tipo de "lente", pero realmente es más bien un filtro o guía de luz, que hace que la luz baje por la fibra delgada donde, probablemente, se produce la absorción. Al otro extremo está la fibra nerviosa. La fibra central está rodeada en sus lados por seis células que, de hecho, han secretado la fibra. Esta descripción es suficiente para nuestros propósitos, lo importante es que es algo cónico y muchas de ellas pueden encajar una al lado de la otra en toda la superficie del ojo de la abeja.

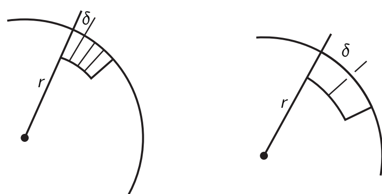


Figura 36-8. Vista esquemática de la disposición de los omatidios en el ojo de una abeja.

Discutamos ahora la resolución del ojo de la abeja. Si dibujamos líneas (figura 36-8) para representar los omatidios en la superficie, que suponemos que es una esfera de radio r, podemos *calcular* realmente el ancho de cada omatidio usando nuestro cerebro ¡y suponiendo que la evolución es tan inteligente como nosotros! Si tenemos un omatidio muy grande, no tenemos mucha resolución. O sea, una célula obtiene una parte de información desde una dirección y la célula adyacente obtiene una parte de información de otra dirección y así sucesivamente, y la abeja no puede ver muy bien las cosas que están por el medio. Así, la indeterminación en la agudeza visual del ojo va a corresponder, seguramente, a un ángulo, el ángulo del extremo del omatidio respecto al centro de curvatura del ojo. (Las células del ojo existen, por supuesto, sólo en la superficie de la esfera; dentro de ésta está la cabeza de la abeja.) Este ángulo desde un omatidio al siguiente es, por supuesto, el diámetro del omatidio dividido entre el radio de la superficie del ojo:

$$\Delta\theta_g = \delta/r. \tag{36.1}$$

De manera que podemos decir "Mientras más delgado hagamos δ, mayor será la agudeza visual. Entonces, ¿por qué la abeja no usa un omatidio muy, pero muy fino?" *Respuesta*: sabemos suficiente física para darnos cuenta de que si queremos hacer pasar luz por una ranura angosta, no podemos ver con precisión en una dirección dada, debido al efecto de difracción. La luz que viene de diferentes

* El ojo humano también tiene una leve sensibilidad a la polarización de la luz ¡y uno puede aprender a ubicar la dirección del Sol! El fenómeno implicado se llama *cepillo de Haidinger*; es una figura muy débil amarillenta, con forma de vidrio de reloj, que se ve en el centro del campo visual, cuando uno observa una gran extensión sin rasgos distintivos, usando anteojos polarizantes. También se puede ver en el cielo azul, sin los anteojos de polarización, si uno rota la cabeza de uno a otro lado alrededor del eje de visión.

† Evidencia obtenida después que esta clase fue dictada, indica que el ojo es sensible directamente.

direcciones puede entrar y, debido a la difracción, vamos a tener luz entrando en un ángulo $\Delta\theta_d$, tal que

$$\Delta\theta_d = \lambda/\delta. \tag{36.2}$$

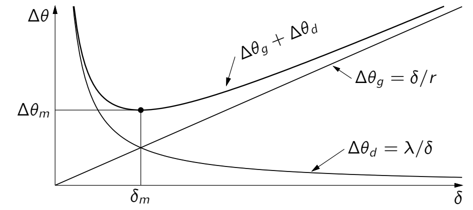


Figura 36-9. El tamaño óptimo de un omatidio es δ_m.

Ahora vemos que si hacemos δ demasiado pequeño, ¡cada omatidio no ve solamente en una dirección, debido a la difracción! Si los hacemos demasiado grandes, cada uno ve en una dirección definida, pero no hay suficiente número de ellos para tener una buena visión de la escena. Por lo tanto, ajustamos la distancia δ para hacer mínimo el efecto total de estos dos. Si los sumamos y buscamos el lugar donde la suma tiene un mínimo (figura 36-9) encontramos que:

$$\frac{d(\Delta\theta_g + \Delta\theta_d)}{d\delta} = 0 = \frac{1}{r} - \frac{\lambda}{\delta^2}, \tag{36.3}$$

lo que nos da una distancia

$$\delta = \sqrt{\lambda r}. \tag{36.4}$$

Si suponemos que r mide alrededor de 3 milímetros, tomamos la luz que ve la abeja como de 4 000 ángstroms, juntamos ambas y extraemos la raíz cuadrada, obtenemos:

$$\begin{aligned} \delta &= (3 \times 10^{-3} \times 4 \times 10^{-7})^{1/2}\ \text{m} \\ &= 3.5 \times 10^{-5}\ \text{m} = 35\ \mu. \end{aligned} \tag{36.5}$$

El libro dice que el diámetro es $30\,\mu$ ¡así que esto es una concordancia bastante buena! Así que, aparentemente, en realidad funciona ¡y podemos entender qué determina el tamaño del ojo de la abeja! También es fácil volver a introducir los números y encontrar de qué calidad realmente es la resolución angular del ojo de la abeja; es muy mala, comparada con la de nosotros. Podemos ver cosas que son 30 veces más chicas, en tamaño aparente, que la abeja; la abeja tiene una imagen borrosa fuera de foco, comparado con lo que nosotros podemos ver. Sin embargo, está bien, y es lo más que pueden hacer. Podríamos preguntarnos por qué las abejas no desarrollan un ojo bueno como el nuestro con una lente, etc. Hay varias razones interesantes. En primer lugar, la abeja es demasiado chica; si tuviera un ojo como el nuestro, pero a su escala, la abertura sería del orden de $30\,\mu$ y la difracción sería tan importante que no podría ver muy bien de todos modos. El ojo no sirve si es muy chico. En segundo lugar, si fuera tan grande como la cabeza de la abeja, el ojo ocuparía toda la cabeza de la abeja. Lo hermoso del ojo compuesto es que no ocupa lugar, es sólo una delgada capa en la superficie de la abeja. De manera que cuando alegamos que deberían haberlo hecho de nuestra manera, ¡debemos recordar que ellas tenían sus propios problemas!

36-5 Otros ojos

Además de las abejas, muchos animales pueden ver color. Los peces, las mariposas, los pájaros y los reptiles pueden ver color, pero se cree que la mayoría de los mamíferos no pueden. Los primates pueden ver color. Los pájaros por cierto que ven color y eso explica los colores de los pájaros. ¡No tendría ningún sentido tener machos con colores tan brillantes, si las hembras no pudieran verlo! Esto es, la evolución de lo sexual "lo que sea que esto signifique", lo que tienen los pájaros es el resultado de la capacidad de la hembra de ver el color. De manera que la próxima vez que veamos un pavo real y pensemos en el brillante despliegue de colores maravillosos que es y en lo delicados que son los colores y qué maravilloso sentido estético se necesita para apreciar todo esto, no debemos felicitar al pavo real, sino ensalzar la agudeza visual y sentido estético de la *hembra* ¡porque eso es lo que ha generado tan bella escena!

Todos los invertebrados tienen ojos pobremente desarrollados u ojos compuestos, pero todos los vertebrados tienen ojos muy similares a los nuestros, con una excepción. Si consideramos la forma más evolucionada de animales,

decimos generalmente. "¡Aquí estamos!", pero si tomamos un punto de vista de menos prejuicios y nos restringimos a los invertebrados, de modo que no podamos incluirnos, y preguntamos cuál es el animal invertebrado más evolucionado, la mayoría de los zoólogos están de acuerdo que ¡el *pulpo* es el animal más desarrollado! Es muy interesante que, además del desarrollo de su cerebro y sus reacciones, etc., que son bastante buenas para un invertebrado, ha desarrollado, en forma independiente, un ojo diferente. No es un ojo compuesto ni es un ojo puntual —tiene córnea, párpados, iris, una lente, tiene dos regiones de agua y tiene una retina atrás—. ¡Es esencialmente igual al ojo de los vertebrados! Es un notable ejemplo de coincidencia en la evolución, donde la naturaleza ha descubierto dos veces la misma solución para un problema, con un pequeño mejoramiento. Resulta también sorprendente que en el pulpo la retina es un pedazo de cerebro, que ha salido de la misma manera en su desarrollo de embrión que en los vertebrados, pero lo interesante y diferente es que las células que son sensibles a la luz están en el *interior* y las células que hacen los cálculos están detrás de aquéllas, en vez de "al revés", como en nuestro ojo. Así vemos que por lo menos no hay ninguna buena razón para que esté al revés. ¡La otra vez que la naturaleza lo ensayó, lo puso al derecho! (Ver figura 36-10). Los ojos más grandes del mundo son los del calamar gigante, ¡se han encontrado de hasta cerca de 40 centímetros de diámetro!

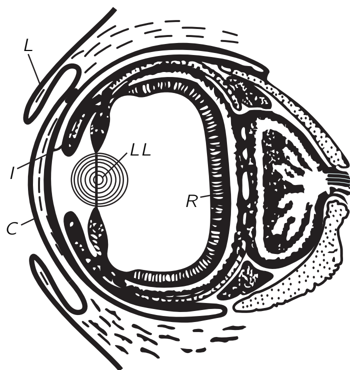
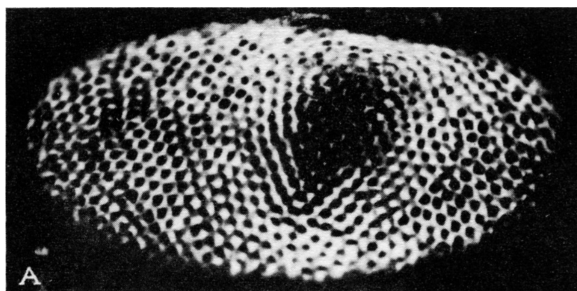


Figura 36-10. El ojo de un pulpo

36-6 Neurología de la visión

Uno de los puntos más importantes de nuestro tema es la interconexión de la información de una parte del ojo a la otra. Consideremos el ojo compuesto del cangrejo gigante límulo, sobre el cual se ha hecho gran cantidad de experimentos. Primero, tenemos que apreciar qué tipo de información puede venir a lo largo de los nervios. Un nervio lleva un tipo de perturbación que tiene un efecto eléctrico fácil de detectar, un tipo de perturbación ondulatoria que baja por el nervio y produce un efecto en el otro extremo: un largo pedazo de la célula nerviosa, llamado axón, lleva la información y un cierto impulso, llamado "espique"*, se propaga si es excitado en un extremo. Cuando un espique baja por un nervio, otro no puede seguir inmediatamente. Todos los espiques son del mismo tamaño, de manera que no tenemos espiques *más altos* cuando la cosa se excita más fuertemente, sino que obtenemos *más espiques por segundo*. El *tamaño* del espique queda determinado por la fibra. Es importante darse cuenta de esto para ver qué sucede a continuación.

La figura 36-11(a) muestra el ojo compuesto del límulo; no tiene mucho de ojo, tiene solamente cerca de 1 000 omatidios. La figura 36-11(b) es una sección transversal del sistema; se puede ver el omatidio con las fibras nerviosas que salen de ellos y van al cerebro. Pero nótese que aun en un límulo hay pequeñas



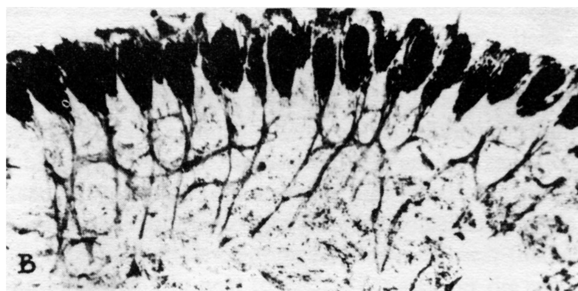


Figura 36-11. El ojo compuesto del cangrejo gigante límulo. (a) Vista normal. (b) Sección transversal.

* N. del T. La palabra inglesa «spike», para designar un pulso rápido de muy corta duración, es también de vasto uso en la terminología castellana.

interconexiones. Son mucho menos elaboradas que el ojo humano y nos dan la oportunidad de estudiar un ejemplo más simple.

Veamos ahora los experimentos que se han hecho poniendo finos electrodos en el nervio óptico de un límulo y haciendo incidir luz en un solo omatidio, lo que es fácil de hacer con lentes. Si encendemos la luz en un instante t_0 y medimos los pulsos eléctricos que salen, encontramos que hay un pequeño retraso y luego una rápida serie de descargas que gradualmente disminuye a una rapidez uniforme, como se muestra en la figura 36-12(a). Cuando la luz se apaga, la descarga se detiene. Ahora bien, es muy interesante que si, mientras nuestro amplificador está conectado a esta misma fibra nerviosa, iluminamos un omatidio *diferente*, no sucede nada; no hay señal.

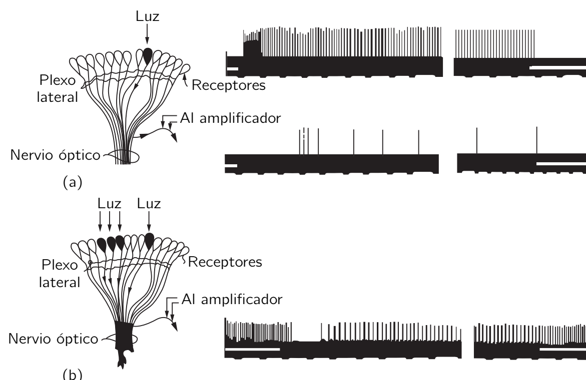


Figura 36-12. La respuesta a la luz de las fibras nerviosas del ojo del límulo.

Ahora hacemos otro experimento: iluminamos el omatidio original y obtenemos la misma respuesta, pero si iluminamos otro cercano también, los pulsos se interrumpen brevemente y después se suceden a una rapidez mucho menor (figura 36-12b). ¡La rapidez de una es inhibida por los impulsos que vienen del otro! En otras palabras, cada fibra nerviosa lleva la información desde un omatidio, pero la cantidad que lleva es inhibida por las señales de los otros. De manera que si, por ejemplo, todo el ojo está iluminado en forma más o menos uniforme, la información proveniente de cualquiera de los omatidios va a ser relativamente débil, porque está inhibida por muchos otros. De hecho, la inhibición es aditiva —si iluminamos varios omatidios muy cercanos, la inhibición es muy grande—. La inhibición es mayor cuando los omatidios están más juntos, y si los omatidios están lo suficientemente alejados entre sí, la inhibición es prácticamente cero. De manera que es aditiva y depende de la distancia; aquí tenemos un primer ejemplo de cómo la información proveniente de diferentes partes del ojo se combina en el ojo mismo. Podemos ver, quizá, si pensamos en eso un rato, que es un dispositivo para *destacar los contrastes* en los bordes de los objetos, porque si una parte de la escena está iluminada y una parte es negra, entonces los omatidios en el área iluminada producen impulsos que son inhibidos por la otra luz en la vecindad, de manera que es relativamente débil. Por otro lado, un omatidio en el borde, al cual se le da un impulso "blanco", también es inhibido por otros en la vecindad, pero no existen tantos, ya que algunos son negros; la señal resultante es, por lo tanto, más fuerte. El resultado sería una curva, algo así como la de la figura 36-13. El crustáceo va a percibir un realce del borde.

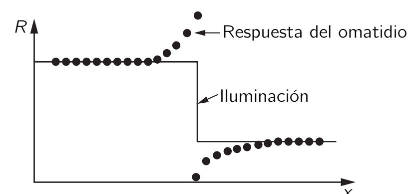


Figura 36-13. La respuesta neta de los omatidios del límulo cerca de un cambio brusco de iluminación.

El hecho de que exista un realce de los bordes se conoce desde hace mucho; en realidad, es un hecho notable que ha sido comentado muchas veces por los psicólogos. Para dibujar un objeto, basta que dibujemos un esbozo. ¡Cómo estamos acostumbrados a mirar cuadros que tienen sólo el esbozo! ¿Qué es el esbozo? El esbozo es sólo la diferencia de contorno entre luz y sombra o un color y otro. No es algo definido. No es, créanlo o no, que cada objeto tenga una línea alrededor de él. No existe tal línea. Es sólo en nuestra propia construcción psicológica donde hay una línea, empezamos a entender la razón por la cual la

"línea" es clave suficiente para llegar al todo. Probablemente nuestros propios ojos trabajan en forma similar —mucho más complicada, pero similar—.

Finalmente, describiremos brevemente el trabajo más detallado, el trabajo hermoso y avanzado realizado en el sapo. Haciendo un experimento correspondiente en un sapo, colocando en el nervio óptico del sapo unas sondas en forma de agujas muy finas y construidas muy bellamente, uno puede obtener las señales que van a lo largo de un axón particular y, lo mismo que en el caso del cangrejo gigante, encontramos que la información no depende de sólo un punto del ojo, sino que es la suma de la información de varios puntos.

El cuadro más reciente de la manera en que funciona el ojo del sapo es el siguiente. Uno puede encontrar cuatro tipos diferentes de fibras nerviosas ópticas, en el sentido de que hay cuatro tipos de respuestas distintas. Estos experimentos no se hicieron iluminando con pulsos de luz, porque eso no es lo que ve un sapo. El sapo se queda sentado simplemente y sus ojos no se mueven nunca, a menos que la hoja de lirio se bambolee de un lado a otro, y en ese caso sus ojos oscilan en forma tal, que la imagen queda fija. Él no mueve sus ojos. Si algo se mueve en su campo visual, como un pequeño insecto (tiene que ser capaz de ver algo moviéndose en un fondo fijo), resulta que hay cuatro tipos diferentes de fibras que descargan, cuyas propiedades se resumen en la tabla 36-1. La detección sostenida del borde, imborrable, significa que si introducimos un objeto con un borde dentro del campo de visión del sapo, entonces hay muchos impulsos en esta fibra particular, mientras el objeto se está moviendo, pero disminuyen, hasta ser un impulso sostenido, que permanece mientras el borde esté ahí, aun si está quieto. Si apagamos la luz, los impulsos se detienen. Si la prendemos otra vez mientras el borde aún está a la vista, empiezan de nuevo. No son borrables. Otro tipo de fibra es muy similar, salvo que si el borde es recto, no funciona. ¡Debe ser un borde convexo con fondo oscuro! ¡Cómo debe de ser de complicado el sistema de interconexiones en la retina del ojo del sapo, para entender que una superficie convexa ha aparecido! Además, aunque esta fibra mantiene algo, no mantiene tanto como la otra y si apagamos la luz y la volvemos a encender, *no* vuelve a producirse el impulso. Depende del movimiento de la superficie convexa. El ojo lo ve entrar y recuerda que está ahí, pero si simplemente apagamos la luz por un momento, sencillamente se olvida y no lo ve más.

Otro ejemplo es la detección del cambio de contraste. Si hay un borde entrando o saliendo, hay pulsos; pero si el objeto se mantiene inmóvil, no hay pulsos.

También existe un detector de penumbra. Si la intensidad de la luz baja, produce pulsos, pero si permanece baja o aumenta, los pulsos se detienen, funciona sólo cuando la luz se está debilitando.

Finalmente, también existen unas cuantas fibras, que son detectores de oscuridad —algo de lo más sorprendente—, ¡disparan todo el tiempo! Si aumentamos la luz, disparan menos rápido, pero todo el tiempo. Si disminuimos la luz, disparan más rápido, todo el tiempo. En la oscuridad, disparan como locos, diciendo en forma perpetua: "¡Está oscuro! ¡Está oscuro! ¡Está oscuro!"

Ahora bien, estas respuestas son bastante difíciles de clasificar y podríamos pensar que a lo mejor estos experimentos fueron mal interpretados. ¡Sin embargo,

Tabla 36-1

Tipos de respuestas en fibras nerviosas ópticas de un sapo

Tipo	Velocidad	Campo angular
1. Detección sostenida del borde (imborrable)	0.2-0.5 m/s	1°
2. Detección del borde convexo (borrable)	0.5 m/s	2°-3°
3. Detección de contraste cambiante	1-2 m/s	7°-10°
4. Detección de penumbra	Hasta $\frac{1}{2}$ m/s	Hasta 15°
5. Detección de oscuridad	?	Muy grande

es muy interesante que estas mismas clases estén separadas muy claramente en la anatomía del sapo! Mediante otras mediciones, después de que estas respuestas fueron clasificadas (*después*, esto es lo importante), se descubrió que la velocidad de las señales en las diferentes fibras no era la misma, ¡de modo que aquí tenemos otra manera independiente de comprobar qué tipo de fibra hemos encontrado!

Otra pregunta interesante es: ¿desde qué tamaño de área hace una fibra particular sus cálculos? La respuesta es diferente para distintas clases.

La figura 36-14 muestra la superficie del así llamado tectum de un sapo, donde los nervios entran al cerebro desde el nervio óptico. Todas las fibras nerviosas que llegan del nervio óptico se conectan en diversas capas del tectum. Esta estructura en capas es análoga a la retina; esto es, en parte, la razón por la que sabemos que el cerebro y la retina son muy similares. Ahora bien, tomando un electrodo e introduciéndolo sucesivamente a través de las capas, podemos averiguar dónde termina cada tipo de nervio óptico, ¡y el resultado, hermoso y maravilloso, es que los diferentes tipos de fibras terminan en capas diferentes! Las primeras terminan en el tipo número 1, las segundas en el número 2, las terceras y quintas acaban en el mismo lugar y la más profunda de todas es la número 4. (¡Qué coincidencia, tienen los números casi en el orden correcto! No, ésa es la razón por la cual las enumeraron de esa manera, ¡la primera publicación tenía los números en diferente orden!)

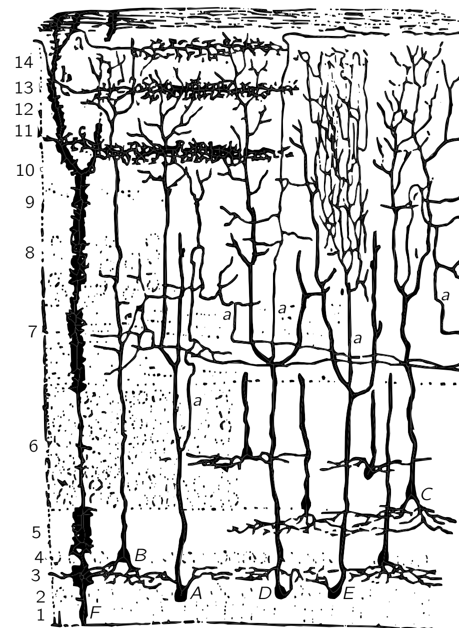


Figura 36-14. El tectum de un sapo.

Podemos resumir brevemente lo que hemos aprendido, de esta manera: probablemente hay tres pigmentos. Puede haber muchos tipos diferentes de células receptoras que contienen los tres pigmentos en distintas proporciones, pero hay muchas interconexiones que permiten sumas y restas mediante adiciones y refuerzos en el sistema nervioso. De manera que antes de que entendamos la visión de los colores, tendremos que entender la sensación final. Este tema está aún abierto, pero estas investigaciones con microelectrodos y otros instrumentos, a lo mejor nos darán, en último término, mayor información sobre cómo vemos el color.

BIBLIOGRAFÍA

Committee on Colorimetry, Optical Society of America, *The Science of Color*, Thomas Y. Crowell Company, Nueva York, 1953.

"Mechanisms of Vision", segundo suplemento del *Journal of General Physiology*, vol. 43, núm. 6, parte 2, julio de 1960, Rockefeller Institute Press.

ARTÍCULOS ESPECÍFICOS

De Robertis, E., "Some Observations on the Ultrastructure and Morphogenesis of Photoreceptors", pp. 1-15.

Hurvich, L. M., y D. Jameson, "Perceived Color, Induction Effects, and Opponent-Response Mechanisms", pp. 63-80.

Rosenblith, W. A., ed., *Sensory Communication*, Massachusetts Institute of Technology Press, Cambridge, Mass., 1961.

"Sight, Sense of", *Encyclopaedia Britannica*, vol. 20, 1957, pp. 628-635.

37

Comportamiento cuántico

37-1 Mecánica atómica

En los últimos capítulos hemos tratado las ideas esenciales necesarias para la comprensión de la mayoría de los fenómenos importantes de la luz —o, en general, de la radiación electromagnética—. (Hemos dejado algunos tópicos especiales para el próximo año. Específicamente, la teoría de los índices de los materiales densos y la reflexión total interna.) En lo que estamos interesados es en la llamada “teoría clásica” de las ondas eléctricas, la cual llega a ser una adecuada y completa descripción de la naturaleza para un gran número de efectos. Todavía no nos hemos preocupado del hecho que la luz llega en gránulos o “fotones”.

Nos gustaría considerar como nuestro siguiente tema el problema del comportamiento de relativamente grandes pedazos de materia —por ejemplo, sus propiedades mecánicas y térmicas—. En su discusión, encontraremos que la teoría “clásica” (o antigua) falla casi inmediatamente, ya que la materia está realmente constituida de partículas de tamaño atómico. Sin embargo, trataremos solamente la parte clásica, ya que es la única parte que podemos entender, usando la mecánica clásica que hemos estado aprendiendo. Pero no tendremos mucho éxito. Encontraremos que en el caso de la materia, a diferencia del caso de la luz, nos encontraremos relativamente pronto, en dificultades. Podríamos, por supuesto, evadir continuamente los efectos atómicos, pero en vez de ello interpondremos aquí una corta excursión, en la cual describiremos las ideas básicas de las propiedades cuánticas de la materia, es decir, las ideas cuánticas de la física atómica, de modo que ustedes se den cuenta de aquello que dejamos de lado. Porque tendremos que dejar de lado algunos temas importantes que no podemos evitar que se nos presenten.

De modo que daremos ahora la *introducción* al tema de la mecánica cuántica, pero no podremos entrar realmente en el tema hasta más adelante.

La “mecánica cuántica” es la descripción del comportamiento de la materia en sus detalles y en particular de lo que sucede a escala atómica. A una escala muy pequeña, las cosas se comportan de un modo distinto a aquello en lo que ustedes tengan alguna experiencia directa. No se comportan como ondas, tampoco como partículas, no se comportan como nubes, o bolas de billar, o pesos sobre resortes, o como nada que hayan visto alguna vez.

Newton pensaba que la luz estaba constituida de partículas, pero entonces se descubrió, como hemos visto aquí, que se comporta como una onda. Más tarde, sin embargo (en los comienzos del siglo XX), se descubrió que en realidad la luz se comporta algunas veces como una partícula. Históricamente, se pensó que el electrón, por ejemplo, se comportaba como una partícula, y entonces se encontró que en muchos aspectos se comportaba como una onda. Así que realmente no se comporta ni como lo uno ni como lo otro. Ahora nos damos por vencidos. Decimos: “No es *ni lo uno ni lo otro*”.

Sin embargo, hay un hecho afortunado: los electrones se comportan exactamente como la luz. El comportamiento cuántico de objetos atómicos (electrones, protones, neutrones, fotones, etc.) es el mismo para todos, todos ellos son “ondas corpusculares”, o como quieran llamarlas. De modo que lo que aprendamos respecto a las propiedades de los electrones (los cuales usaremos para nuestros ejemplos) se aplicará a todas las “partículas” incluyendo los fotones de la luz.

La acumulación gradual de información sobre el comportamiento atómico y a pequeña escala durante el primer cuarto del siglo XX, lo cual dio algunas indicaciones de cómo se comportan los objetos pequeños, produjo una confusión

creciente, que finalmente se resolvió en 1926 y 1927 por Schrödinger, Heisenberg y Born. Ellos, finalmente, obtuvieron una descripción sistemática del comportamiento de la materia a escala pequeña. En el presente capítulo trataremos los principales aspectos de esa descripción.

Como el comportamiento atómico es tan diferente a la experiencia común, es muy difícil acostumbrarse y a todos aparece como algo peculiar y misterioso, tanto al novato como al físico experimentado. Aun los expertos no lo entienden en la forma en que les gustaría hacerlo, y esto es perfectamente razonable, ya que toda la experiencia y la intuición humanas se aplican a objetos grandes. Nosotros sabemos cómo actuarán estos objetos grandes, pero cosas a escala pequeña no actúan precisamente en esa forma. De modo que tenemos que aprender respecto a ellas en una cierta forma abstracta o imaginativa y no en conexión con nuestra experiencia directa.

En este capítulo abordaremos en forma inmediata el elemento básico del misterioso comportamiento en su más extraña forma. Elijamos examinar un fenómeno, que es imposible, *absolutamente* imposible, de explicar en cualquier forma clásica, y que contiene la esencia de la mecánica cuántica. En realidad, contiene el *único* misterio. No podemos explicar el misterio en el sentido de una "explicación" de cómo funciona. Sólo les *diremos* cómo funciona. Diciéndoles cómo funciona esto, les habremos dicho las particularidades básicas de toda la mecánica cuántica.

37-2 Un experimento con proyectiles

Para tratar de entender el comportamiento cuántico de los electrones, compararemos y confrontaremos su comportamiento, en un dispositivo experimental particular, con el comportamiento más familiar de partículas, como proyectiles, y con el comportamiento de ondas, como las ondas del agua. En primer lugar, consideraremos el comportamiento de los proyectiles en el dispositivo experimental mostrado esquemáticamente en la figura 37-1. Tenemos una ametralladora que dispara un chorro de proyectiles. No es un arma muy buena, ya que esparce los proyectiles (al azar) sobre un intervalo angular bastante amplio, como se indica en la figura. Frente al cañón tenemos una pared (hecha de lámina blindada) que tiene dos agujeros suficientemente grandes para dejar pasar un proyectil. Más allá de la pared hay una contención (digamos una gruesa pared de madera), la cual "absorbe" los proyectiles cuando éstos la golpean. Frente a la pared tenemos un objeto que llamaremos "detector" de proyectiles. Éste puede ser una caja que contenga arena. Cualquier proyectil que penetre al detector será detenido y acumulado. Cuando lo deseemos podemos vaciar la caja y contar el número de proyectiles que han sido atrapados. El detector puede moverse hacia adelante y hacia atrás (la que llamaremos dirección x). Con este dispositivo podemos encontrar experimentalmente la respuesta a la pregunta: "¿Cuál es la probabilidad de que un proyectil que atraviese los agujeros en la pared llegue a la contención a una distancia x del centro?" En primer lugar, deben darse cuenta de que tenemos que hablar de probabilidad, ya que no podemos decir en forma

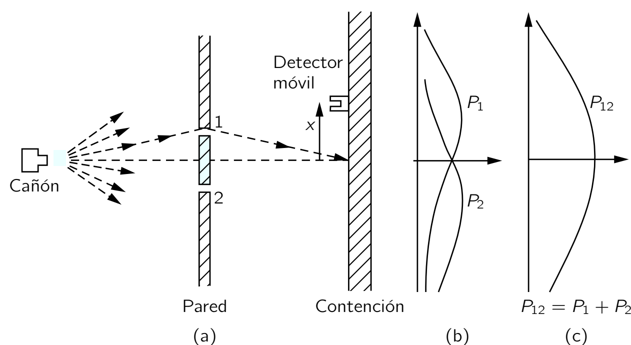


Figura 37-1. Experimento de interferencia con proyectiles.

definitiva a dónde irá un proyectil particular. Un proyectil, que por casualidad golpea uno de los agujeros, puede rebotar en los bordes del agujero y puede terminar en cualquier parte. Por "probabilidad" queremos decir la posibilidad de que el proyectil llegue al detector, que podemos medir contando el número que llega al detector en un cierto tiempo, y luego haciendo el cociente entre este número y el número *total* que golpea la contención durante ese tiempo. O bien, si suponemos que el arma siempre dispara con la misma rapidez durante las medidas, la probabilidad que queremos es exactamente proporcional al número que alcanza el detector durante algún intervalo patrón de tiempo.

Para nuestros propósitos actuales nos gustaría imaginar algún experimento idealizado en el cual los proyectiles no sean proyectiles reales, sino proyectiles *indestructibles* —que no puedan partirse en mitades—. En nuestro experimento encontramos que los proyectiles siempre llegan en gránulos, y cuando encontramos algo en el detector, siempre es un proyectil entero. Si la rapidez con la cual la ametralladora dispara se hace muy baja, encontramos que en cualquier momento dado no llega ningún proyectil a la contención, o bien llega uno y solamente uno —exactamente uno—. También el tamaño de los gránulos ciertamente no depende de la rapidez de disparo del arma. Diremos: "Los proyectiles *siempre* llegan en gránulos idénticos". Lo que medimos con nuestro detector es la probabilidad de llegada de un gránulo y medimos la probabilidad en función de x. El resultado de tales medidas con este aparato (no hemos realizado todavía el experimento, así que realmente estamos imaginando el resultado) está trazado en la gráfica dibujada en la parte (c) de la figura 37-1. En la gráfica dibujamos la probabilidad hacia la derecha y verticalmente x, de modo que la escala de x se ajusta al diagrama del dispositivo. Llamamos P_{12} a la probabilidad porque los proyectiles pueden haber venido ya sea a través del agujero 1 o bien a través del agujero 2. Ustedes no se sorprenderán de que P_{12} sea grande cerca del medio del gráfico, pero que se haga chico si x es muy grande. Sin embargo, podrían preguntarse por qué P_{12} tiene su máximo valor en $x = 0$. Podemos entender este hecho, si hacemos otra vez nuestro experimento, después de cubrir el agujero 2, y una vez más mientras se cubre el agujero 1. Cuando el agujero 2 está cubierto, los proyectiles pueden pasar solamente a través del agujero 1 y obtenemos la curva indicada por P_1 en la parte (b) de la figura. Como podrían esperar, el máximo de P_1 ocurre en el valor de x que está en línea recta con el cañón y el agujero 1. Cuando el agujero 1 está cerrado, obtenemos la curva simétrica P_2, trazada en la figura. P_2 es la distribución de probabilidad para los proyectiles que pasan a través del agujero 2. Comparando las partes (b) y (c) de la figura 37-1, encontramos el importante resultado que:

$$P_{12} = P_1 + P_2. \tag{37.1}$$

Las probabilidades justamente se suman. El efecto con ambos agujeros abiertos es la suma de los efectos con cada agujero abierto separadamente. Llamaremos este resultado una observación de "*no interferencia*": por una razón que ustedes más tarde verán. Hasta aquí los proyectiles. Llegan en gránulos y su probabilidad de llegada no muestra interferencia.

37-3 Un experimento con ondas

Ahora deseamos considerar un experimento con ondas en agua. El aparato se muestra esquemáticamente en la figura 37-2. Tenemos un recipiente poco profundo con agua. Un pequeño objeto indicado por "fuente de ondas" está vibrando hacia arriba y hacia abajo mediante un motor y produce ondas circulares. A la derecha de la fuente tenemos otra vez una pared con dos agujeros y más allá de ella hay una segunda pared, la cual, para mantener simples las cosas, es un "absorbente", de modo que no hay reflexión de las ondas que llegan allí. Esto puede hacerse construyendo una "playa" gradual de arena. Frente a la playa colocamos un detector, el que como antes se puede mover para adelante y para atrás, en la dirección x. El detector es ahora un dispositivo que mide la

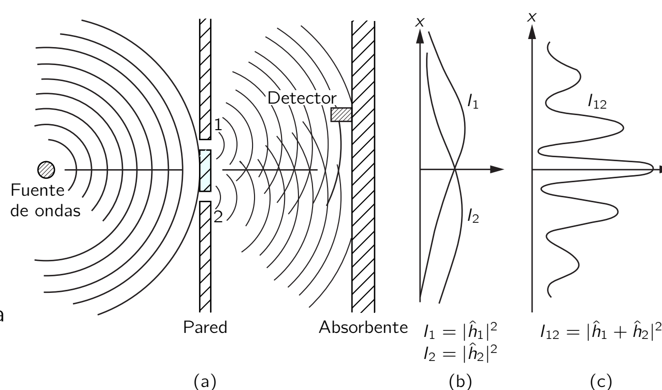


Figura 37-2. Experimento de interferencia con ondas de agua.

"intensidad" del movimiento ondulatorio. Pueden imaginar un dispositivo que mide la altura del movimiento ondulatorio, pero cuya escala está calibrada en proporción al *cuadrado* de la altura real, de modo que la lectura es proporcional a la intensidad de la onda. Entonces nuestro detector lee en proporción a la *energía* transportada por la onda —o más bien, la cantidad de energía que llega al detector por unidad de tiempo—.

Con nuestro aparato de ondas, la primera cosa que se observa es que la intensidad puede tener *cualquier* tamaño. Si la fuente mueve una muy pequeña cantidad, hay sólo un poco de movimiento ondulatorio en el detector. Cuando hay más movimiento en la fuente, más intensidad hay en el detector. La intensidad de la onda puede tener cualquier valor. *No* diríamos que hay alguna "granulación" en la intensidad de la onda.

Ahora midamos la intensidad de la onda para varios valores x (cuidando que la fuente de ondas funcione siempre en la misma forma). Obtenemos la curva de aspecto interesante indicada con I_{12} en la parte (c) de la figura.

Hemos calculado ya cómo tales figuras pueden producirse, cuando estudiamos la interferencia de ondas eléctricas. En este caso observamos que la onda original se difracta en los agujeros, y nuevas ondas circulares se propagan desde cada agujero. Si cubrimos un agujero un instante y medimos la distribución de intensidad en el absorbente, encontramos las curvas de intensidad bastante simples, mostradas en la parte (b) de la figura. I_1 es la intensidad de la onda proveniente del agujero 1 (que encontramos haciendo la medición cuando el agujero 2 está bloqueado) e I_2 es la intensidad de la onda proveniente del agujero 2 (vista cuando el agujero 1 está bloqueado).

La intensidad observada I_{12} cuando ambos agujeros están abiertos ciertamente *no* es la suma de I_1 e I_2. Decimos que hay "interferencia" de las dos ondas. En algunos lugares (donde la curva I_{12} tiene sus máximos) las ondas están "en fase" y los máximos de las ondas se suman para dar una gran amplitud y, por lo tanto, una gran intensidad. Decimos que las dos ondas están "interfiriendo constructivamente" en tales lugares. Habrá tal interferencia constructiva siempre que la distancia desde el detector a uno de los agujeros sea mayor (o menor), en un número entero de longitudes de onda, que la distancia desde el detector al otro agujero.

En aquellos lugares donde las dos ondas llegan al detector con un desfase de π (donde están "fuera de fase"), el movimiento ondulatorio resultante en el detector será la diferencia de las dos amplitudes. Las ondas "interfieren destructivamente", y obtenemos un valor bajo para la intensidad de onda. Esperamos tales valores dondequiera que la distancia entre el agujero 1 y el detector difiere de la distancia del agujero 2 al detector en un número impar de cuasilongitudes de ondas. Los valores bajos de I_{12} en la figura 37-2 corresponden a los lugares donde las dos ondas interfieren destructivamente.

Recordarán que la relación cuantitativa entre I_1, I_2 e I_{12} puede expresarse en la siguiente forma: la altura instantánea de la onda de agua en el detector puede ser escrita para la onda proveniente del agujero 1 como (la parte real

de) $\hat{h}_1 e^{i\omega t}$, donde la “amplitud” $\hat{h}_1$, en general, es un número complejo. La intensidad es proporcional al cuadrado medio de la altura, o cuando usamos los números complejos, a $|\hat{h}_1|^2$. Similarmente para el agujero 2, la altura es $\hat{h}_2 e^{i\omega t}$ y la intensidad es proporcional a $|\hat{h}_2|^2$. Cuando ambos agujeros están abiertos, las alturas de las ondas se suman para dar la altura $(\hat{h}_1 + \hat{h}_2)e^{i\omega t}$ y la intensidad $|\hat{h}_1 + \hat{h}_2|^2$. Omitiendo la constante de proporcionalidad para nuestros propósitos presentes, las relaciones apropiadas para *ondas que interfieren* son:

$$I_1 = |\hat{h}_1|^2, \qquad I_2 = |\hat{h}_2|^2, \qquad I_{12} = |\hat{h}_1 + \hat{h}_2|^2. \tag{37.2}$$

Notarán que el resultado es bastante diferente del obtenido con proyectiles (ecuación 37.1). Si desarrollamos $|\hat{h}_1 + \hat{h}_2|^2$ vemos que:

$$|\hat{h}_1 + \hat{h}_2|^2 = |\hat{h}_1|^2 + |\hat{h}_2|^2 + 2|\hat{h}_1||\hat{h}_2| \cos\delta, \tag{37.3}$$

donde δ es el desfase entre $\hat{h}_1$ y $\hat{h}_2$. En función de las intensidades, podríamos escribir:

$$I_{12} = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\delta. \tag{37.4}$$

El último término en (37.4) es el “término de interferencia”. Suficiente con las ondas de aguas. La intensidad puede tener cualquier valor y muestra interferencia.

37-4 Un experimento con electrones

Imaginemos ahora un experimento similar, pero con electrones. Se muestra esquemáticamente en la figura 37-3. Construyamos un cañón de electrones que consiste en un alambre de tungsteno calentado mediante una corriente eléctrica y rodeado por una caja de metal con un agujero en ella. Si el alambre está a un voltaje negativo respecto a la caja, los electrones emitidos por el alambre serán acelerados hacia las paredes y algunos pasarán a través del agujero. Todos los electrones que saldrán del cañón tendrán (aproximadamente) la misma energía. Frente al cañón se tiene otra vez una pared (simplemente una delgada lámina metálica) con dos agujeros en ella. Más allá de la pared hay otra lámina que servirá como una “contención”. Frente a la contención colocamos un detector móvil. El detector puede ser un contador Geiger o, quizá mejor, un multiplicador de electrones, conectado a un altoparlante.

Deberíamos decirles inmediatamente que no traten de disponer dicho experimento (como pudieron haberlo hecho con los dos ya descritos). Este experimento nunca ha sido hecho en esta forma. El problema es que el aparato tendría que ser construido a una escala imposiblemente pequeña para mostrar los efectos en los que estamos interesados. Estamos haciendo un “experimento pensado”, el cual hemos elegido porque es fácil pensar en él. Sabemos los resultados que *se obtendrían* ya que *hay* muchos experimentos que han sido hechos, en los cuales la escala y las proporciones han sido elegidas para mostrar los efectos que describiremos.

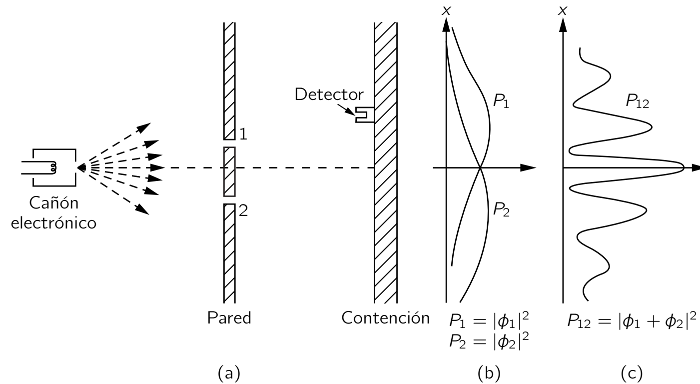


Figura 37-3. Experimento de interferencia con electrones.

La primera cosa que notamos con nuestro experimento de electrones es que oímos marcados "clics" provenientes del detector (esto es, del hablante). Y todos los "clics" son iguales. *No* hay "medios clics".

También deberíamos notar que los "clics" llegan muy erráticamente. Algo como: clic . . . . . . clic-clic . . . clic . . . . . . clic . . . clic-clic . . . . . . clic . . . . . . etc., como ustedes, sin duda, han oído funcionar un contador Geiger. Si contamos los clics que llegan en un tiempo suficientemente largo —muchos minutos, digamos— y luego contamos otra vez durante otro periodo igual, encontramos que los dos números son aproximadamente iguales. De modo que podemos hablar de una *frecuencia media* a la cual los clics se escuchan (tantos clics por minuto en promedio).

A medida que movemos el detector, la *frecuencia* a la cual los clics aparecen es mayor o menor, pero el tamaño (sonoridad) de cada clic es siempre el mismo. Si bajamos la temperatura del alambre en el cañón, la frecuencia de los clics baja, pero todavía cada clic suena igual. También podríamos notar que si ponemos dos detectores separados sobre la contención, uno *u* otro sonaría, pero nunca ambos al mismo tiempo. (Excepto que en algún momento, si hubiera dos clics muy cerca en el tiempo, nuestro oído no pudiera sentir la separación.) Por lo tanto, concluimos que dondequiera que lleguen a la contención, llegan en "gránulos". Todos los "gránulos" son del mismo tamaño, solamente llegan "gránulos" enteros, y llegan uno a uno a la contención. Diremos: "Los electrones llegan siempre en gránulos idénticos".

Al igual que en nuestro experimento con proyectiles, podemos ahora proseguir para encontrar experimentalmente la respuesta a la pregunta: "¿Cuál es la probabilidad relativa de que un 'gránulo' electrónico llegue a la contención a diversas distancias x del centro?" Como antes, obtenemos la probabilidad relativa, observando la frecuencia de los clics, manteniendo constante el funcionamiento del cañón. La probabilidad con la cual los "gránulos" llegan a un x particular es proporcional a la frecuencia media de los clics en ese x.

El resultado de nuestro experimento es la interesante curva, marcada P_{12} en la parte (c) de la figura 37-3. ¡Sí! Ésa es la forma en que se comportan los electrones.

37-5 La interferencia de ondas de electrones

Tratemos ahora de analizar la curva de la figura 37-3 para ver si podemos entender el comportamiento de los electrones. Lo primero que deberíamos decir es que, ya que ellos llegan en gránulos, cada gránulo, que podemos llamar igualmente bien un electrón, ha pasado, o a través del agujero 1, o bien del agujero 2. Escribamos esto en la forma de "proposición":

Proposición A: Cada electrón pasa, ya sea a través del agujero 1 *o bien* a través del agujero 2.

Suponiendo la proposición A, todos los electrones que lleguen a la contención pueden ser divididos en dos clases: (1) los que llegan a través del agujero 1 y (2) los que a través del agujero 2. De modo que nuestra curva observada debe ser la suma de los efectos de los electrones que pasan a través del agujero 1 y los electrones que pasan a través del agujero 2. Verifiquemos esta idea mediante un experimento. Primero, realizaremos una medida para los electrones que pasan a través del agujero 1. Taparemos el agujero 2 y realizaremos nuestras cuentas de los clics del detector. De la frecuencia de los clics obtenemos P_1. Así, el resultado de la medida se muestra mediante la curva marcada P_1 en la parte (b) de la figura 37-3. El resultado parece bastante razonable. En una forma similar, medimos P_2, la distribución de probabilidad para los electrones que pasan a través del agujero 2. El resultado de esta medida está también trazado en la figura.

El resultado P_{12} obtenido con *ambos* agujeros abiertos claramente no es la suma de P_1 y P_2, las probabilidades para cada agujero por separado. Por analogía con nuestro experimento de ondas de agua, decimos: "hay interferencia".

$$\textit{Para electrones:} \qquad P_{12} \neq P_1 + P_2. \tag{37.5}$$

¿Cómo pudo producirse esa interferencia? Quizá podríamos decir: "Bien, eso significa presumiblemente que *no es verdad* que los gránulos pasan, ya sea a través del agujero 1 o del agujero 2, porque si lo hicieran, las probabilidades deberían sumarse. Quizá van en una forma más complicada. Se separan en mitades y. . . " ¡Pero no! No pueden, siempre llegan en gránulos. . . "Bien, quizás algunos de ellos pasen a través de 1 y luego den la vuelta a través del 2, y luego den vueltas algunas veces más, o mediante algún otro camino complicado. . . entonces cerrando el agujero 2 cambiaríamos la probabilidad de que un electrón que *partiera* a través del agujero 1 llegara finalmente a la contención. . . " ¡Pero observen! Hay algunos puntos a los cuales llegan muy pocos electrones, cuando *ambos* agujeros están abiertos, pero que reciben muchos electrones si cerramos un agujero, de modo que *cerrando* un agujero *aumenta* el número proveniente desde el otro. Sin embargo, noten que en el centro de la figura P_{12} es más del doble de grande que $P_1 + P_2$. Se ha pensado que cerrando un agujero *disminuiría* el número de electrones que pasan a través del otro agujero. Parece difícil de explicar *ambos* efectos, al proponer que los electrones viajen según complicadas trayectorias.

Todo es bastante misterioso, y mientras más lo miran, más misterioso parece. Muchas ideas han sido inventadas para tratar de explicar la curva P_{12} en función de los electrones individuales que pasan a través de los agujeros en formas complicadas. Ninguna de ellas ha tenido éxito. Ninguna de ellas ha podido obtener la curva correcta para P_{12} en función de P_1 y P_2.

Sin embargo, es bastante sorprendente que la *matemática* para relacionar P_1 y P_2 con P_{12} sea extremadamente simple. Porque P_{12} es precisamente como la curva I_{12} de la figura 37-2 y *ésa* era simple. Lo que sucede en la contención puede describirse con dos números complejos que podemos llamar $\hat{\phi}_1$ y $\hat{\phi}_2$ (son funciones de x, por supuesto). El módulo al cuadrado de $\hat{\phi}_1$ da el efecto con sólo el agujero 1 abierto. Esto es, $P_1 = |\hat{\phi}_1|^2$. El efecto con sólo el agujero 2 abierto está dado por $\hat{\phi}_2$ en la misma forma. Esto es $P_2 = |\hat{\phi}_2|^2$. Y el efecto combinado de los dos agujeros es justamente $P_{12} = |\hat{\phi}_1 + \hat{\phi}_2|^2$. ¡La *matemática* es la misma que la que teníamos para las ondas de agua! (Es difícil ver cómo se pudo obtener un resultado tan simple de un complicado juego de electrones que van y vienen a través de la lámina con alguna extraña trayectoria.)

Concluimos lo siguiente: los electrones llegan en gránulos como partículas y la probabilidad de llegada de esos gránulos está distribuida como la distribución de la intensidad de una onda. En este sentido es que los electrones se comportan "algunas veces como una partícula y algunas veces como una onda".

A propósito, cuando estábamos tratando ondas clásicas, definimos la intensidad como el promedio temporal del cuadrado de la amplitud de la onda y usamos números complejos como un artificio matemático para simplificar el análisis. Pero en mecánica cuántica resulta que las amplitudes *deben* ser representadas por números complejos. Sólo las partes reales no servirán. Por el momento, esto es un punto técnico, ya que las fórmulas se ven exactamente iguales.

Ya que la probabilidad de llegada a través de ambos agujeros está dada tan simplemente, a pesar de que no es igual a $(P_1 + P_2)$, eso es realmente todo lo que hay que decir. Pero hay un gran número de sutilezas implicadas en el hecho de que la naturaleza funciona en esa forma. Nos gustaría ilustrar ahora algunas de esas sutilezas para ustedes. Primero, como el número que llega a un punto particular no es igual al número que llega a través de 1 más el número que llega a través de 2, como podríamos haber concluido de la proposición A, indudablemente concluiríamos que la *proposición A es falsa. No* es verdad que los electrones van *bien* a través del agujero 1 o bien del agujero 2. Pero tal conclusión puede ser comprobada mediante otro experimento.

37-6 Observando los electrones

Ensayemos ahora el siguiente experimento. Agreguemos a nuestro aparato electrónico una fuente de luz muy intensa, colocada detrás de la pared y entre los dos agujeros, como se muestra en la figura 37-4. Sabemos que las cargas eléctricas dispersan la luz. Así que cuando un electrón pasa, siempre que pase,

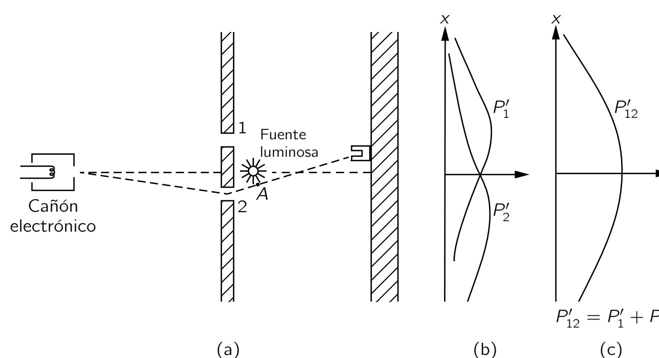


Fig. 37-4. Un experimento diferente con electrones.

en su camino al detector, dispersará algo de luz hacia nuestro ojo, y podremos *ver* dónde va el electrón. Si, por ejemplo, un electrón tomara la trayectoria vía agujero 2 esquematizada en la figura 37-4, deberíamos ver un destello de luz proveniente de la vecindad del lugar marcado A en la figura. Si un electrón pasa a través del agujero 1 deberíamos esperar ver un destello desde la vecindad del agujero superior. Si sucediera que obtenemos luz desde ambos lugares al mismo tiempo porque el electrón se divide en mitades... ¡Mejor hagamos el experimento!

He aquí lo que vemos: *cada* vez que escuchamos un "clic" de nuestro detector de electrones (en la contención), *también vemos* un destello de luz *o bien* cerca del agujero 1 o cerca del agujero 2, ¡pero *nunca* de ambos a la vez! Y observamos el mismo resultado sin que importe dónde colocamos el detector. De esta observación concluimos que, si observamos electrones, encontramos que los electrones pasan o bien a través de un agujero o del otro. Experimentalmente la proposición A es necesariamente verdadera.

Entonces, ¿qué está errado en nuestro argumento *en contra* de la proposición A? ¿Por qué P_{12} *no es* simplemente igual a $P_1 + P_2$? ¡De nuevo al experimento! Sigámosles la pista a los electrones y averigüemos qué están haciendo. Para cada posición (posición x) del detector contaremos los electrones que llegan y *también* tomaremos nota a través de cuál agujero pasaron, observando los destellos. Podemos seguir el curso de las cosas de esta manera: siempre que escuchamos un "clic", colocaremos una cuenta en la columna 1, si vemos un destello cerca del agujero 1, y si vemos un destello cerca del agujero 2 registraremos una cuenta en la columna 2. Cada electrón que llega se registra en una de las dos clases: los que pasan a través de 1 y los que pasan a través de 2. A partir del número registrado en la columna 1, obtenemos la probabilidad P'_1 de que un electrón llegue al detector vía agujero 1, y a partir del número registrado en la columna 2 obtenemos P'_2, la probabilidad de que un electrón llegue al detector vía agujero 2. Si repetimos ahora tal medida, para muchos valores de x, obtenemos las curvas para P'_1 y P'_2 que se muestran en la parte (b) de la figura 37-4.

¡Bien, esto no es tan sorprendente! Obtenemos para P'_1 algo muy similar a lo que obtuvimos antes para P_1 bloqueando el agujero 2; y P'_2 es similar a lo que obtuvimos, bloqueando el agujero 1. Así que *no* existe ningún asunto complicado como el de pasar a través de ambos agujeros. Cuando los observamos, los electrones atraviesan exactamente como habríamos esperado de ellos que pasaran. Sea que los agujeros estén cerrados o abiertos, los que hemos visto pasar a través del agujero 1 están distribuidos en la misma forma, sea que el agujero 2 esté abierto o cerrado.

¡Pero esperen! ¿Qué tenemos *ahora* para la probabilidad *total*, la probabilidad de que un electrón llegue al detector por cualquier ruta? Ya tenemos esta información. Simplemente pretendemos que no observamos nunca los destellos luminosos y juntamos los clics del detector que hemos separado en dos columnas. *Debemos sumar* simplemente los números. Para la probabilidad de que un electrón llegue a la contención, pasando a través de *cualquier* agujero, encontramos $P'_{12} = P'_1 + P'_2$. Esto es, a pesar de que pudimos ver a través de cuál agujero pasan nuestros electrones, no obtenemos más la antigua curva de interferencia

P_{12} sino que una nueva, P'_{12}, ¡que no muestra interferencia! Si apagamos la luz, se restablece la P_{12}.

Debemos concluir que *cuando observamos los electrones*, su distribución sobre la pantalla es diferente que cuando no los observamos. ¿Es quizás el encendido de nuestra fuente luminosa el que perturba las cosas? Debe ser que los electrones son muy delicados y cuando la luz dispersa los electrones, les da una sacudida que cambia sus movimientos. Sabemos que el campo eléctrico de la luz al actuar sobre una carga ejercerá una fuerza sobre ella. Así, pues, *deberíamos* esperar quizá que el movimiento cambie. Por esto, la luz ejerce una gran influencia sobre los electrones. Al tratar de "mirar" los electrones, hemos cambiado sus movimientos. Así, la sacudida dada al electrón cuando el fotón es dispersado por aquél es tal que cambia el movimiento del electrón lo suficiente, de modo que si hubiera *podido* ir hacia donde P_{12} era un máximo; irá, en cambio, donde P_{12} era un mínimo; así es que no vemos más los efectos ondulatorios de interferencia.

Puede que estén pensando: "¡No use una fuente tan brillante! ¡Baje la intensidad! Entonces las ondas luminosas serán más débiles y no perturbarán tanto los electrones. Seguramente, haciendo la luz más y más mortecina, eventualmente la onda será lo suficientemente débil de modo que tendrá un efecto despreciable". De acuerdo, ensayémoslo. Lo primero que observamos es que los destellos de luz dispersados por los electrones a medida que pasan *no* se hacen más débiles. *Es siempre un destello de la misma intensidad.* Lo único que sucede cuando la luz se hace más débil es que algunas veces escuchamos un "clic" del detector, pero no vemos *ningún destello.* El electrón ha pasado sin ser "visto". Lo que estamos observando es que la luz *también* actúa como los electrones, *sabíamos* que era "ondulosa", pero ahora encontramos que también es "granulosa". Siempre llega —o es dispersada— en gránulos que llamamos "fotones". Cuando bajamos la *intensidad* de la fuente luminosa, no cambiamos el *tamaño* de los fotones, solamente la *frecuencia* a la cual se emiten. *Eso* explica por qué, cuando nuestra fuente es débil, algunos electrones pasan sin ser vistos. No dio la casualidad de que hubiera un fotón cerca en el instante en que el electrón pasara.

Todo esto es un poco desalentador. Si es cierto que siempre que "vemos" el electrón, vemos un destello de la misma intensidad, entonces los electrones que vemos son *siempre* los perturbados. Ensayemos de todos modos el experimento con una luz débil. Ahora, siempre que escuchemos un clic en el detector, registraremos una cuenta en tres columnas: en la columna (1) los electrones vistos por el agujero 1, en la columna (2) los electrones vistos por el agujero 2 y en la columna (3) los electrones que no se han visto en absoluto. Cuando elaboramos nuestros datos (calculando las probabilidades) encontramos estos resultados: los "vistos por el agujero 1" tienen una distribución como P'_1; los "vistos por el agujero 2" tienen una distribución como P'_2 (de modo que los "vistos tanto por el agujero 1 como por el 2" tienen una distribución como P_{12}), ¡y los "no vistos en absoluto" tienen una distribución "ondulosa" como P_{12} de la figura 37-3! *¡Si los electrones no se ven, tenemos interferencia!*

Eso es comprensible. Cuando no vemos el electrón, ningún fotón lo perturba y, si lo vemos, un fotón lo ha perturbado. Siempre existe el mismo monto de perturbación porque los fotones de luz producen todos efectos de la misma intensidad y el efecto de los fotones dispersados es suficiente para borrar cualquier efecto de interferencia.

¿No existirá *alguna* manera de que podamos ver los electrones sin perturbarlos? Aprendimos en un capítulo anterior que el momentum transportado por un "fotón" es inversamente proporcional a su longitud de onda ($p = h/\lambda$). Ciertamente, la sacudida dada al electrón cuando el fotón es dispersado hacia nuestro ojo depende del momentum que transporta el fotón. ¡Ah! Si queremos perturbar los electrones sólo ligeramente, no deberíamos haber bajado la *intensidad* de la luz, deberíamos haber bajado su *frecuencia* (lo mismo que aumentar su longitud de onda). Usemos luz de un color más rojo. Podríamos usar aún luz infrarroja u ondas de radio (como el radar), y ver dónde fue el electrón con la ayuda de algún equipo que pueda "ver" luz de estas longitudes de onda más larga. Si usamos luz "más suave" podemos quizás evitar de perturbar tanto los electrones.

Ensayemos el experimento con ondas más largas. Continuaremos repitiendo nuestro experimento cada vez con luz de una longitud de onda más larga. Al comienzo, nada parece cambiar. Los resultados son los mismos. Entonces sucede una cosa terrible. Ustedes recordarán que cuando discutimos el microscopio recalcamos que, debido a la *naturaleza ondulatoria* de la luz, existe una limitación en cuanto a lo cerca que pueden estar dos manchas y todavía ser vistas como dos manchas separadas. Esta distancia es del orden de la longitud de onda de la luz. Así ahora, cuando hacemos la longitud de onda mayor que la distancia entre nuestros agujeros vemos un destello *grande* borroso cuando la luz es dispersada por los electrones. ¡No podemos decir más a través de cuál agujero pasó el electrón! ¡Sólo sabemos que se fue a alguna parte! Y es precisamente con luz de este color que encontramos que las sacudidas dadas al electrón son lo suficientemente pequeñas para que P'_{12} empiece a parecerse a P_{12} —que empezamos a obtener algún efecto de interferencia—. Y es solamente para longitudes de onda mucho más grande que la separación de los dos agujeros (cuando no tenemos posibilidad alguna de decir por dónde pasó el electrón) que la perturbación debida a la luz se hace lo suficientemente pequeña como para que obtengamos de nuevo la curva P_{12} mostrada en la figura 37-3.

En nuestro experimento encontramos que es imposible disponer la luz de manera tal que se pueda decir por cuál agujero pasó el electrón y al mismo tiempo no perturbar el diagrama. Heisenberg sugirió que las entonces nuevas leyes de la naturaleza sólo podían ser compatibles si existiera alguna limitación básica en nuestra capacidad experimental, no reconocida previamente. Propuso como un principio general su *principio de incertidumbre*, que podemos enunciar en términos de nuestro experimento como sigue: “Es imposible diseñar un dispositivo para determinar a través de cuál agujero pasa el electrón, que no perturbe al mismo tiempo los electrones lo suficiente para destruir el diagrama de interferencia”. Si un dispositivo es capaz de determinar a través de cuál agujero pasa el electrón, *no puede* ser tan delicado que no perturbe el diagrama de una manera esencial. Nadie ha encontrado jamás (ni aun imaginado) una manera de obviar el principio de incertidumbre. Así, pues, debemos suponer que describe una característica básica de la naturaleza.

La teoría completa de la mecánica cuántica que usamos ahora para describir los átomos y, de hecho, toda la materia, depende de la validez del principio de incertidumbre. Dado que la mecánica cuántica es una teoría de tanto éxito, nuestra confianza en el principio de incertidumbre se refuerza. Pero si se descubriera alguna vez una manera de “abatir” el principio de incertidumbre, la mecánica cuántica daría resultados incompatibles y debería descartarse como teoría válida de la naturaleza.

“Bien”, dirán ustedes, “¿qué hay con la proposición A? ¿Es verdad o *no* es verdad, que el electrón pase, ya sea por el agujero 1 o por el agujero 2?” La única respuesta que se puede dar es que hemos encontrado a partir del experimento que hay cierta manera especial en que debemos pensar para no caer en contradicciones. Lo que debemos decir (para evitar hacer predicciones erradas) es lo siguiente. Si uno observa los agujeros, o más precisamente, si uno tiene un aparato que sea capaz de determinar si los electrones pasan por el agujero 1 o por el agujero 2, uno *puede* decir que pasa, ya sea por el agujero 1 o por el agujero 2. *Pero* si uno *no* trata de decir qué camino toman los electrones, cuando no hay nada en el experimento para perturbar los electrones, entonces uno *no* puede decir que un electrón pase, ya sea por el agujero 1 o por el agujero 2. Si uno dice eso, y comienza a hacer deducciones cualesquiera a partir de esta afirmación, hará errores en el análisis. Éste es el filo lógico por el cual debemos caminar si deseamos describir con éxito la naturaleza.

Si el movimiento de toda la materia —como también de los electrones— debe ser descrito en términos de ondas, ¿qué hay con los proyectiles en nuestro primer experimento? ¿Por qué no vimos allí un diagrama de interferencia? Resulta que para proyectiles las longitudes de onda eran tan diminutas que los diagramas de

interferencia se hacían muy finos. Tan finos, en efecto, que con cualquier detector de tamaño distinto de cero uno no podría distinguir los máximos y mínimos separados. Lo que vimos era sólo una especie de promedio, que es la curva clásica. En la figura 37-5 tratamos de indicar esquemáticamente qué sucede con objetos a escala grande. La parte (a) de la figura muestra la distribución de probabilidad que uno podría predecir para proyectiles, usando la mecánica cuántica. Las rápidas oscilaciones se suponen representar el diagrama de interferencia que se obtiene para ondas de longitud de onda muy corta. Cualquier detector físico, sin embargo, se monta sobre varias oscilaciones de la curva de probabilidad, tal que las mediciones muestran la curva suave dibujada en la parte (b) de la figura.

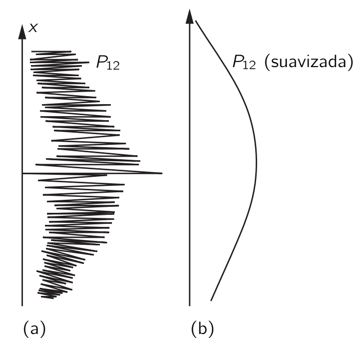


Figura 37-5. Figura de interferencia con proyectiles: (a) real (esquemática), (b) observada.

37-7 Primeros principios de la mecánica cuántica

Escribiremos ahora un resumen de las principales conclusiones de nuestros experimentos. Sin embargo, pondremos los resultados en una forma que los haga válidos para un tipo general de tales experimentos. Podemos escribir nuestro resumen en la forma más simple, si primero definimos un "experimento ideal" como aquel en el cual no hay influencias externas inciertas, es decir, no suceden pequeñas agitaciones u otras cosas que nosotros no podamos tomar en cuenta. Seríamos completamente precisos si dijéramos: "Un experimento ideal es aquel en el cual todas las condiciones iniciales y finales del experimento están completamente especificadas". Lo que llamaremos "un evento" es, en general, un conjunto específico de condiciones iniciales y finales. (Por ejemplo, "un electrón deja el cañón, llega al detector, y nada más sucede".) Ahora nuestro resumen.

RESUMEN

(1) La probabilidad de un evento en un experimento ideal está dada por el cuadrado del valor absoluto de un número complejo ϕ el cual se llama Amplitudes de probabilidad.

$$\begin{aligned} P &= \text{probabilidad}, \\ \phi &= \text{amplitud de probabilidad}, \\ P &= |\phi|^2. \end{aligned} \tag{37.6}$$

(2) Cuando un evento ocurre en varias formas alternativas, la amplitud de probabilidad del evento es la suma de las amplitudes de probabilidad para cada uno considerado separadamente. Hay interferencia.

$$\begin{aligned} \phi &= \phi_1 + \phi_2, \\ P &= |\phi_1 + \phi_2|^2. \end{aligned} \tag{37.7}$$

(3) Si se ejecuta un experimento que sea capaz de determinar si una u otra alternativa es la que realmente ocurrió, la probabilidad del evento es la suma de las probabilidades de cada alternativa. La interferencia se pierde.

$$P = P_1 + P_2. \tag{37.8}$$

A uno todavía le gustaría preguntar: "¿Cómo funciona esto? ¿Cuál es el mecanismo detrás de la ley?" Nadie ha encontrado ningún mecanismo detrás de la ley. Nadie puede "explicar" nada más de lo que ya hemos "explicado". Nadie les dará una representación más profunda de la situación. No tenemos idea acerca de un mecanismo más básico del que puedan deducirse estos resultados.

Nos gustaría poner énfasis en una diferencia muy importante entre mecánica clásica y cuántica. Hemos estado hablando sobre la probabilidad de que un electrón llegue en una circunstancia dada. Hemos supuesto que en nuestro dispositivo experimental (o aun en el mejor posible) debería ser imposible predecir exactamente qué va a suceder. ¡Solamente podemos predecir las probabilidades!

Esto significa que, de ser verdad, la física ha desistido del problema de tratar de predecir exactamente lo que sucederá en una determinada circunstancia. ¡Sí! La física *ha* desistido. *No sabemos cómo predecir qué sucedería en una circunstancia dada*, y ahora creemos que esto es imposible, que lo único que puede ser predicho es la probabilidad de diferentes eventos. Debe reconocerse que esto es una disminución en nuestro primitivo ideal en la comprensión de la naturaleza. Puede ser un paso atrás, pero nadie ha encontrado un modo de evitarlo.

Ahora haremos algunas observaciones respecto a una sugerencia que se ha hecho algunas veces para tratar de evitar la descripción que hemos dado: "Quizás el electrón tiene algún tipo de capacidad interna —algunas variables internas— que no conocemos aún. Es por esto, tal vez que no podemos predecir qué sucederá. Si pudiéramos observar de más cerca el electrón, podríamos estar en situación de decir a dónde irá a parar". Hasta donde sabemos, esto es imposible. Aún así tendríamos dificultades. Supongan que hiciéramos la hipótesis de que en el interior del electrón hay cierto tipo de mecanismo que determina a dónde irá a parar. Ese mecanismo debe *también* determinar a través de cuál agujero pasará en su camino. Pero no debemos olvidar que lo que está dentro del electrón no debería depender de lo que *nosotros* hacemos y, en particular, de si abrimos o cerramos uno de los agujeros. Por lo tanto, si un electrón antes de partir tiene ya en mente (a) cuál agujero va a usar y (b) dónde irá a parar, encontraríamos P_1 para los electrones que han elegido el agujero 1, P_2 para los que han elegido el agujero 2, *y necesariamente* la suma $P_1 + P_2$ para los que llegan a través de los dos agujeros. Parece que no hay otro camino para obviar esto. Pero hemos verificado experimentalmente que éste no es el caso. Y nadie ha encontrado una salida a este enigma. De manera que actualmente debemos limitarnos a calcular probabilidades. Decimos "actualmente", pero sospechamos fuertemente, que es algo que estará con nosotros para siempre —que es imposible resolver ese enigma— que ésta es la manera como la naturaleza realmente *es*.

37-8 El principio de incertidumbre

Ésta es la forma en que Heisenberg originalmente enunció el principio de incertidumbre: si hacen la medida sobre cualquier objeto, y pueden determinar la componente x de su momentum con una incertidumbre Δp, no pueden al mismo tiempo conocer su posición x en forma más precisa que $\Delta x \geq \hbar/2\Delta p$. Las incertidumbres en la posición y el momentum en cualquier instante deben tener su producto mayor o igual que la mitad de la constante de Planck reducida. Éste es un caso especial del principio de incertidumbre que fue enunciado anteriormente con más generalidad. El enunciado más general fue de que ningún modo uno puede diseñar un equipo que determine cuál de las dos alternativas debe tomarse, sin que al mismo tiempo se destruya el diagrama de interferencia.

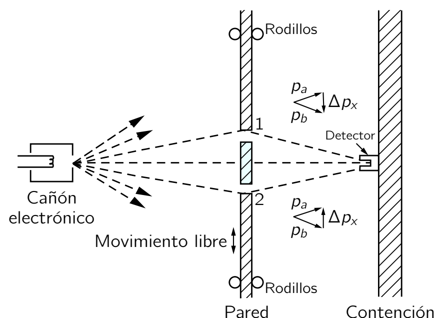


Figura 37-6. Un experimento en que se mide el retroceso de la pared.

Demostremos, para un caso particular, que el tipo de relación dado por Heisenberg debe ser verdadero, para evitar tener problemas. Imaginemos una modificación del experimento de la figura 37-3, en la cual la pared con los agujeros consiste en una lámina montada sobre rodillos de modo que pueda moverse libremente hacia arriba y hacia abajo (en la dirección x), como se muestra en la figura 37-6. Observando cuidadosamente el movimiento de la lámina, podemos tratar de decir a través de cuál agujero pasa un electrón. Imaginemos lo que sucede cuando el detector está colocado en $x = 0$. Esperaríamos que un electrón que pasa por el agujero 1 debe ser desviado hacia abajo por la lámina para poder alcanzar el detector. Como la componente vertical del momentum del electrón cambia, la lámina debe retroceder, con un momentum igual en la dirección opuesta. La lámina recibirá un empuje hacia arriba. Si el electrón pasa a través del agujero inferior, la lámina debería experimentar un empuje hacia abajo. Está claro que para cada posición del detector, el momentum recibido por la lámina tendrá un valor diferente para una travesía vía agujero 1 que para una travesía vía agujero 2. ¡Así es! Sin perturbar en *ninguna forma* los electrones, sino sólo observando la *lámina*, podemos decir qué trayectoria usó el electrón.

Ahora bien, para hacer esto es necesario conocer el momentum de la lámina, antes de que el electrón la atraviese. De manera que cuando medimos el momentum después que el electrón haya pasado, podemos calcular cuánto ha cambiado el momentum de la lámina. Pero recuerden, según el principio de incertidumbre, no podemos conocer al mismo tiempo la posición de la lámina, con una precisión arbitraria. Pero si no sabemos exactamente *dónde* está la lámina, no podemos decir con precisión dónde están los dos agujeros. Estarán en un lugar diferente para cada electrón que atraviesa. Esto quiere decir que el centro de nuestra figura de interferencia tendrá una ubicación diferente para cada electrón. Las vibraciones de las figuras de interferencia se harán más borrosas. En el próximo capítulo demostraremos cuantitativamente que si determinamos el momentum de la lámina con suficiente precisión, para determinar a partir de la medida del retroceso, cuál agujero fue utilizado, entonces la incertidumbre en la posición x de la lámina será, de acuerdo al principio de incertidumbre, suficiente para desplazar la figura observada en el detector hacia arriba y hacia abajo en la dirección x en aproximadamente la distancia de un máximo a su mínimo más cercano. Tal corrimiento al azar es suficiente para emborronar la figura, de modo que no se observa interferencia.

El principio de incertidumbre "protege" a la mecánica cuántica. Heisenberg reconoció que, si fuera posible medir el momentum y la posición simultáneamente con una precisión mayor, la mecánica cuántica sufriría un colapso. Así que propuso que esto debe ser imposible. Entonces la gente se sentó y trató de imaginar formas de hacerlo, pero nadie pudo imaginar alguna manera de medir la posición y el momentum de alguna cosa —una pantalla, un electrón, una bola de billar, cualquier cosa— con mayor precisión. Así la mecánica cuántica mantiene su peligrosa pero precisa existencia.

38

Relación entre los puntos de vista ondulatorio y corpuscular

38-1 Amplitudes de ondas de probabilidad

En este capítulo discutiremos la relación entre los puntos de vista ondulatorio y corpuscular. Ya sabemos, por el último capítulo, que ni el punto de vista ondulatorio ni el punto de vista corpuscular son correctos. Usualmente hemos tratado de presentar las cosas en forma precisa, o al menos, tan precisas que no tengan que ser cambiadas cuando aprendamos más —¡pueden ser extendidas, pero no serán cambiadas!— Pero cuando tratamos de hablar sobre la representación ondulatoria o la representación corpuscular, ambas son aproximadas y cambiarán. Por lo tanto, lo que aprendamos en este capítulo no será preciso; en cierto sentido, es una especie de argumento semintuitivo que más tarde se hará más preciso, pero ciertas cosas cambiarán algo cuando las interpretemos correctamente en mecánica cuántica. Por supuesto, la razón de hacer tal cosa es que no estamos entrando directamente a la mecánica cuántica, sino que deseamos tener por lo menos alguna idea de los tipos de efectos que encontraremos. Más aún, todas nuestras experiencias son con ondas y con partículas, y así es bastante práctico usar las ideas ondulatorias y corpusculares para obtener algún conocimiento de lo que sucede en circunstancias dadas, antes de conocer la matemática completa de las amplitudes cuánticas. Trataremos de ilustrar los puntos más débiles a medida que avancemos, pero la mayor parte de esto es casi correcto —es sólo asunto de interpretación—.

Ante todo, sabemos que la nueva forma de representar el mundo en mecánica cuántica —la nueva estructura— es dar una amplitud a cada evento que pueda ocurrir y si el evento involucra la recepción de una partícula, podemos dar la amplitud de encontrar esa partícula en diferentes lugares y en distintos tiempos. La probabilidad de encontrar la partícula es, entonces, proporcional al cuadrado del valor absoluto de la amplitud. En general, la amplitud de encontrar una partícula en diferentes lugares y en distintos tiempos varía con la posición y el tiempo.

En un caso especial, la amplitud varía sinusoidalmente en el espacio y en el tiempo como $e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}$ (no olvidar que estas amplitudes son números complejos, no números reales) y comprende una frecuencia definida ω y un número de onda $\boldsymbol{k}$. Entonces resulta que esto corresponde a la situación límite clásica donde podríamos haber creído tener una partícula cuya energía E fuese conocida y estuviese relacionada a la frecuencia,

$$E = \hbar\omega, \tag{38.1}$$

y cuyo momentum $\boldsymbol{p}$ es también conocido y estuviese relacionado con el número de onda como:

$$\boldsymbol{p} = \hbar\boldsymbol{k}. \tag{38.2}$$

Esto significa que la idea de partícula es limitada. La idea de una partícula —su ubicación, su momentum, etc.— que usamos tanto, es en cierta forma insatisfactoria. Por ejemplo, si una amplitud de encontrar una partícula en lugares diferentes está dada por $e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}$, cuyo módulo al cuadrado es una constante, esto significaría que la probabilidad de encontrar una partícula es la misma en todos los puntos. Esto significa que no sabemos *dónde* está —puede estar en cualquier parte—; hay una gran incertidumbre en su ubicación.

Por otra parte, si la posición de una partícula se conoce más o menos bien podemos predecir esto con bastante precisión, entonces la probabilidad de encontrarla en lugares diferentes debe estar restringida a una cierta región, cuya longitud llamamos Δx. Fuera de esta región, la probabilidad es cero. Ahora bien, esta probabilidad es el cuadrado del módulo de una amplitud, y si el cuadrado del módulo es cero, la amplitud también es cero, de manera que tenemos un tren de ondas cuya longitud es Δx (figura 38-1), y la longitud de onda (la distancia entre nodos de las ondas en el tren) de este tren de ondas es lo que corresponde al momentum de la partícula.

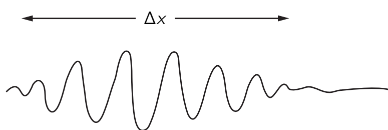


Figura 38-1. Un paquete de ondas cuyo largo es Δx.

Aquí encontramos un hecho extraño respecto a las ondas; algo muy simple, lo cual no tiene nada que ver estrictamente con la mecánica cuántica. Es algo que todos aquellos que trabajan con ondas, aun si no conocen la mecánica cuántica saben, es decir, *no podemos definir una única longitud de onda para un tren de ondas corto.* Tal tren de ondas no tiene una longitud de onda definida, hay una indefinición en el número de onda, la cual está relacionada con la longitud finita del tren y por ello hay una indefinición en el momentum.

38-2 Medida de la posición y del momentum

Consideremos dos ejemplos de esta idea para ver la razón de por qué existe una incertidumbre en la posición y/o en el momentum, siempre que la mecánica cuántica esté correcta. Hemos visto antes también, que si no hubiera tal cosa —si fuera posible medir simultáneamente la posición y el momentum de algo— tendríamos una paradoja, es una fortuna que no tengamos tal paradoja y el hecho de que tal incertidumbre llegue naturalmente a partir de la representación ondulatoria muestra que todo es internamente compatible.

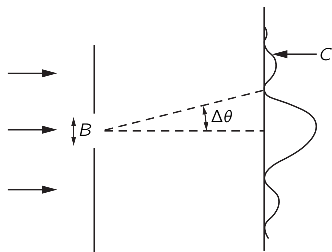


Figura 38-2. Difracción de partículas que pasan a través de una rendija.

Aquí hay un ejemplo que muestra la relación entre la posición y el momentum, en una circunstancia que es fácil de entender. Supongamos tener una sola rendija y partículas que llegan desde muy lejos con una cierta energía —de modo que llegan esencialmente en forma horizontal (figura 38-2)—. Nos concentraremos en las componentes verticales del momentum. Todas estas partículas tienen un cierto momentum horizontal, digamos p_0 en un sentido clásico. De modo que, en el sentido clásico, el momentum vertical p_y, antes de que la partícula pase por el agujero, es perfectamente conocido. La partícula no se mueve ni hacia arriba ni hacia abajo porque viene de una fuente lejana —y así el momentum vertical es, por supuesto, cero—. Pero supongamos ahora que pasa a través del agujero de ancho B. Entonces, después de que ha salido del agujero conocemos la posición en el sentido vertical —la posición y— con considerable precisión, es decir, $\pm B$.* Esto es, la incertidumbre en la posición, Δy, es del orden de B. Ahora quisiéramos también decir, ya que sabemos que el momentum es absolutamente horizontal, que Δp_y es cero; pero esto es erróneo. Nosotros *una vez* supimos que el momentum era horizontal, pero ya no lo sabemos más. Antes de que las partículas pasaran a través del agujero, no conocíamos sus posiciones verticales. ¡Ahora que hemos encontrado la posición vertical, haciendo que la partícula pase a través del agujero, hemos perdido nuestra información sobre el momentum vertical! ¿Por qué? De acuerdo a la teoría ondulatoria, hay un esparcimiento o difracción de las ondas después que pasan la rendija así como para la luz. Por lo tanto, existe una cierta probabilidad de que las partículas que salen de la rendija no sigan exactamente derecho. El diagrama se ensancha por el efecto de difracción y el ángulo de ensanche, que podemos definir como el ángulo del primer mínimo; es una medida de la incertidumbre en el ángulo final.

¿Cómo llega a ensancharse el diagrama? Decir que se ensancha significa que hay alguna probabilidad de que la partícula se esté moviendo hacia arriba o hacia abajo. Esto es, que tenga una componente del momentum hacia arriba o hacia abajo. Decimos *probabilidad* y *partícula* porque podemos detectar este diagrama de difracción con un contador de partículas, y cuando el contador recibe la partícula, digamos en C en la figura 38-2, recibe la partícula *entera*, de modo

* Más precisamente, el error en nuestro conocimiento de y es $\pm B/2$. Pero ahora sólo estamos interesados en la idea general, por lo que no nos preocuparemos por el factor de 2.

que, en un sentido clásico, la partícula tiene un momentum vertical para alcanzar desde la rendija el punto C.

Para obtener una cierta idea de cuánto se esparce el momentum, el momentum vertical p_y se esparce en $p_0\Delta\theta$, donde p_0 es el momentum horizontal. ¿Y de qué tamaño es $\Delta\theta$ en el diagrama ensanchado? Sabemos que el primer mínimo ocurre para un ángulo $\Delta\theta$ tal que las ondas que provienen de un borde de la rendija tienen que viajar una longitud de onda más lejos que las ondas que provienen del otro lado —hemos desarrollado esto antes (capítulo 30)—. Por lo tanto, $\Delta\theta$ es λ/B y así Δp_y en este experimento es $p_0\lambda/B$. Nótese que si hacemos B más pequeño y hacemos una medida más precisa de la posición de la partícula, el diagrama de difracción se ensancha. Recuerden que, cuando cerramos las rendijas en el experimento con las microondas, teníamos más intensidad a ángulos más grandes. De modo que mientras más angosta hacemos la rendija tanto más ancho se hace el diagrama, y mayor es la probabilidad de que encontremos que la partícula tiene momentum hacia el lado. Así pues, la incertidumbre del momentum vertical es inversamente proporcional a la incertidumbre de y. En efecto, vemos que el producto de las dos es igual a $p_0\lambda$. Pero λ es la longitud de onda y p_0 el momentum y, de acuerdo con la mecánica cuántica, la longitud de onda por el momentum es la constante de Planck h. Así, pues, obtenemos la regla que la incertidumbre en el momentum vertical y en la posición vertical tienen un producto del orden de h:

$$\Delta y\,\Delta p_y \geq \hbar/2. \tag{38.3}$$

No podemos preparar un sistema en el cual conozcamos la posición vertical de una partícula y podamos predecir cómo se moverá verticalmente con una certeza mayor que la dada por (38.3). Esto es, la incertidumbre en el momentum vertical debe exceder de $\hbar/2\Delta y$, donde Δy es la incertidumbre en nuestro conocimiento de la posición.

Algunas veces la gente dice que la mecánica cuántica está completamente equivocada. Cuando la partícula llegaba desde la izquierda, su momentum vertical era cero. Y ahora que ha atravesado la rendija, su posición es conocida. Tanto la posición como el momentum parecen conocerse con precisión arbitraria. Es completamente cierto que podemos recibir una partícula y al recibirla determinar cuál es su posición y cuál es el momentum que debería haber tenido para lograr llegar ahí. Esto es verdad, pero eso no es a lo que se refiere la relación de incertidumbre (38.3). La ecuación (38.3) se refiere a la *predictibilidad* de una situación, no se refiere al *pasado*. No está bien decir “Yo sabía cuál era el momentum antes que atravesara la rendija, y ahora conozco su posición”, porque ahora el conocimiento del momentum se ha perdido. El hecho de que haya pasado por la rendija no nos permite predecir el momentum vertical. Estamos hablando respecto a una teoría predictiva, no de medidas, después del hecho, de modo que debemos hablar sobre lo que podemos predecir.

Ahora tomemos el asunto de otra forma. Tomemos otro ejemplo del mismo fenómeno, un poco más cuantitativamente. En el ejemplo anterior medimos el momentum por un método clásico. Es decir, consideramos la dirección, la velocidad, los ángulos, etc., de tal modo que obtuvimos el momentum mediante un análisis clásico. Pero, ya que el momentum está relacionado con el número de onda, todavía existe en la naturaleza otro camino para medir el momentum de una partícula —fotón u otra cosa—, el cual no tiene analogía clásica, ya que utiliza la ecuación (38.2). Medimos la *longitud de onda de las ondulaciones*. Tratemos de medir el momentum en esta forma.

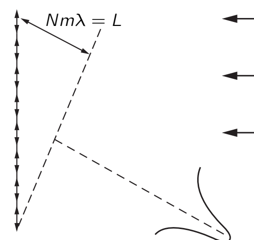


Figura 38-3. Determinación del momentum, usando una red de difracción.

Supongan que tenemos una red de difracción con un gran número de líneas (figura 38-3), y enviamos un haz de partículas hacia la red. Nosotros hemos discutido este problema anteriormente, si las partículas tienen un momentum definido, entonces obtenemos una figura muy marcada en cierta dirección debido a la interferencia. Y también hemos hablado de la precisión con que podemos determinar ese momentum, es decir, cuál es el poder de resolución de esa red. En vez de deducirlo otra vez, nos referimos al capítulo 30, donde encontramos

que la incertidumbre relativa en la longitud de onda que se puede medir con una red dada es $1/Nm$, donde N es el número de líneas de la red y m es el orden del diagrama de difracción. Esto es,

$$\Delta\lambda/\lambda = 1/Nm. \tag{38.4}$$

Ahora bien, la fórmula (38.4) se puede reescribir en la forma:

$$\Delta\lambda/\lambda^2 = 1/Nm\lambda = 1/L, \tag{38.5}$$

donde L es la distancia mostrada en la figura 38-3. Esta distancia es la diferencia entre la distancia total que la partícula u onda o lo que sea tiene que atravesar si se refleja desde la base de la red, y la distancia que tiene que atravesar si es reflejada desde el tope de la red. Es decir, las ondas que forman el diagrama de difracción son ondas que provienen de diferentes partes de la red. Las primeras que llegan provienen de la parte inferior de la red, del comienzo del tren de ondas, y el resto proviene de partes posteriores del tren de ondas, proviniendo de diferentes partes de la red, hasta que la última finalmente llegue, y esto implica un punto en el tren de ondas a una distancia L detrás del primer punto. Así pues, para que tengamos una línea definida en nuestro espectro correspondiente a un momentum definido, con una incertidumbre dada por (38.4), tenemos que contarcon un tren de ondas de una longitud de al menos L. Si el tren de ondas es demasiado corto, no estaremos usando toda la red. Las ondas que forman el espectro se reflejan solamente en un corto sector de la red si el tren de ondas es demasiado corto y la red no trabajará correctamente —encontraremos una gran extensión angular—. Para obtener uno más angosto, necesitamos usar toda la red de modo que al menos en algún momento el tren de ondas completo sea dispersado simultáneamente desde todas las partes de la red. Así el tren de onda debe ser de longitud L para tener una incertidumbre en la longitud de onda menor que la dada por (38.5). A propósito,

$$\Delta\lambda/\lambda^2 = \Delta(1/\lambda) = \Delta k/2\pi. \tag{38.6}$$

Por lo tanto,

$$\Delta k = 2\pi/L. \tag{38.7}$$

donde L es la longitud del tren de ondas.

Esto significa que si tenemos un tren de ondas cuya longitud es menor que L, la incertidumbre en el número de onda debe exceder a $2\pi/L$. O bien, la incertidumbre en el producto de un número de onda por la longitud del tren de ondas —lo llamaremos por un momento Δx— excede 2π. Lo llamaremos Δx porque es la incetidumbre en la posición de la partícula. Si el tren de ondas existe solamente en una longitud finita, entonces es ahí donde podríamos encontrar la partícula dentro de una incertidumbre Δx. Ahora bien, esta propiedad de las ondas de que el producto de la longitud del tren de ondas por la incertidumbre del número de onda asociada con él es por lo menos π, es una propiedad conocida por cualquiera que las haya estudiado. Esto no tiene nada que ver con la mecánica cuántica. Es simplemente, que si tenemos un tren finito no podemos contar las ondas en él en forma muy precisa. Tratemos otra forma de ver la razón de ello.

Supongan que tenemos un tren finito de longitud L; entonces, debido a la forma en que decrece hacia los extremos, como en la figura 38-1, el número de ondas en la longitud L es incierto, en algo como ± 1. Pero el número de ondas en L es $kL/2\pi$. Así, k es incierto y otra vez obtenemos el resultado (38.7), una propiedad nada más que de las ondas. La misma cosa se cumple, ya sea si las ondas están en el espacio y k es el número de radianes por centímetro y L es la longitud del tren, o bien las ondas están en el tiempo y ω es el número de radianes por segundo y T es la "longitud" del tiempo en que llega el tren de ondas. Esto es, si tenemos un tren de onda que dura solamente por un cierto tiempo finito T, la incertidumbre en la frecuencia está dada por:

$$\Delta\omega = 2\pi/T. \tag{38.8}$$

Hemos tratado de recalcar que ésas son las propiedades de las ondas solamente, las cuales son bien conocidas, por ejemplo, en la teoría del sonido.

El punto es que en la mecánica cuántica interpretamos el número de onda como una medida del momentum de una partícula con la regla $p = \hbar k$, de modo que la relación (38.7) nos dice que $\Delta p \approx h/\Delta x$. Entonces, esto es una limitación de la idea clásica del momentum. (¡Naturalmente, esto tiene que estar limitado en alguna forma, si nosotros vamos a representar partículas por ondas!) Es agradable haber encontrado una regla que nos dé alguna idea de cuándo hay fracaso de las ideas clásicas.

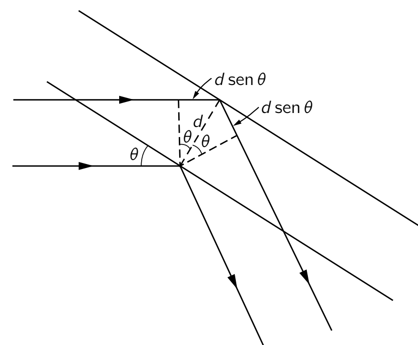


Figura 38-4. Dispersión de ondas por planos cristalinos.

38-3 Difracción en cristales

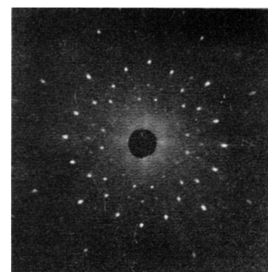
A continuación consideremos la reflexión de ondas corpusculares en un cristal. Un cristal es algo grueso que contiene muchos átomos similares —más tarde incluiremos algunas complicaciones— en un arreglo bonito. La cuestión es cómo ubicar el arreglo de modo que obtengamos una fuerte reflexión máxima en una dirección dada para un haz dado de, digamos, luz (rayos X), electrones, neutrones o cualquier cosa. Para obtener una fuerte reflexión, la dispersión desde todos los átomos debe estar en fase. No puede haber números iguales en fase y fuera de fase, o las ondas se anularán. La forma de arreglar las cosas es encontrar las regiones de fase constante, como ya hemos explicado; son planos que forman ángulos iguales con las direcciones inicial y final (figura 38-4).

Si consideramos dos planos paralelos, como en la figura 38-4, las ondas dispersadas desde los dos planos estarán en fase siempre que la diferencia de distancia recorrida por un frente de onda sea un número entero de longitudes de onda. Esta diferencia se puede ver que es $2d \operatorname{sen} \theta$, donde d es la distancia perpendicular entre los planos. Así pues, la condición para reflexión coherente es

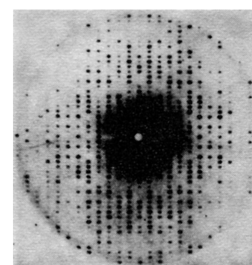
$$2d \operatorname{sen} \theta = n\lambda \qquad (n = 1, 2, \dots). \tag{38.9}$$

Si, por ejemplo, el cristal es tal que los átomos yacen en planos que obedecen la condición (38.9) con $n = 1$, habrá una fuerte reflexión. Si, por otro lado, hay por el medio otros átomos de la misma naturaleza (igual en densidad), entonces los planos intermedios también dispersarán en forma igualmente fuerte e interferirán con los otros y no se producirá efecto alguno. Así, d en (38.9) debe referirse a planos *adyacentes*, ¡no podemos tomar un plano cinco capas más atrás y usar esta fórmula!

Es de interés que los cristales reales usualmente no son tan simples como un solo tipo de átomo repetido en cierta forma. En cambio, si hacemos un análogo en dos dimensiones, son muy similares a una tesela, en el cual hay algún tipo de figura que se repite sobre toda la tesela. Por "figura" entendemos, en el caso de átomos, algún arreglo —calcio, un carbono, tres oxígenos, carbonato de calcio, etc.—, el cual puede comprender un número relativamente grande de átomos. Pero, sea lo que sea, la figura se repite en una estructura. Esta figura básica se llama *celda unitaria*.


Figura 38-5.

La figura básica de repetición define lo que llamamos el *tipo de red cristalina*; el tipo de red se puede inmediatamente determinar examinando las reflexiones y viendo cuál es su simetría. En otras palabras, donde encontramos *alguna* reflexión esto determina el tipo de red, pero para determinar cuáles son cada uno de los elementos de la red, uno debe tomar en cuenta la *intensidad* de la dispersión en varias direcciones. *En qué* direcciones hay dispersión, depende del tipo de red, pero *lo intensa* que es cada dispersión, queda determinado por lo que hay dentro de cada celda unitaria y de esta manera se determina la estructura de los cristales.


Figure 38-6.

En las figuras 38-5 y 38-6 se muestran dos fotografías de diagramas de difracción por rayos X; ilustran la dispersión en sal común y mioglobina, respectivamente.

A propósito, ocurre un hecho interesante si los espacios entre los planos más cercanos son menores que $\lambda/2$. En este caso, la ecuación (38.9) no tiene

solución para n. Así, si λ es mayor que el doble de la distancia entre planos adyacentes, no hay diagrama de difracción lateral y la luz —o lo que sea— atravesará directamente el material sin rebotes o pérdidas. De modo que en el caso de la luz, donde λ es mucho mayor que el espaciado, por supuesto, atraviesa y no hay figura de reflexión en los planos del cristal.

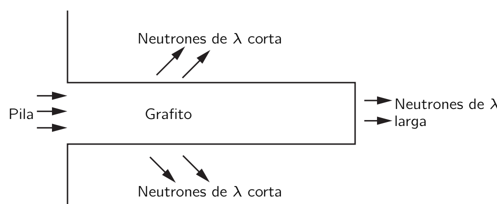


Figura 38-7. Difusión de neutrones de una pila a través de un bloque de grafito.

Este hecho tiene también una consecuencia interesante en el caso de las pilas que producen neutrones (evidentemente éstas son partículas, para cualquier persona). Si tomamos esos neutrones y los introducimos dentro de un gran bloque de grafito, los neutrones se difunden y se abren camino (figura 38-7). Se difunden porque rebotan con los átomos, pero estrictamente, en la teoría ondulatoria, rebotan en los átomos debido a la difracción en los planos cristalinos. ¡Resulta que si tomamos un pedazo muy largo de grafito, los neutrones que salen del extremo más lejano tienen todos una gran longitud de onda! En efecto, si uno hace una gráfica de la intensidad en función de la longitud de onda, no obtenemos nada excepto para longitudes de onda mayores que un cierto mínimo (figura 38-8). En otras palabras, en esa forma podemos obtener neutrones muy lentos. Solamente atraviesan los neutrones más lentos, no son difractados ni dispersados por los planos cristalinos del grafito, sino que continúan atravesando directamente como la luz a través del vidrio, y no son dispersados hacia los lados. Hay muchas otras demostraciones de la realidad de las ondas de neutrones y de las ondas de otras partículas.

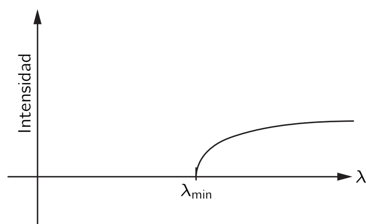


Figura 38-8. Intensidad de neutrones que salen de una barra de grafito, en función de la longitud de onda.

38-4 El tamaño de un átomo

Consideraremos ahora otra aplicación de la relación de incertidumbre, ecuación (38.3). Esto no debe ser tomado demasiado en serio; la idea es correcta, pero el análisis no es muy preciso. La idea tiene que ver con la determinación del tamaño de los átomos y el hecho de que, clásicamente, los electrones irradiaran luz y describirán una espiral hasta depositarse sobre el núcleo. Pero eso no puede ser correcto desde el punto de vista de la mecánica cuántica, porque entonces sabríamos dónde está cada electrón y con qué velocidad se está moviendo.

Supongamos que tenemos un átomo de hidrógeno y que medimos la posición del electrón; no podremos predecir exactamente dónde está el electrón, o de lo contrario, el ancho de la distribución del momentum será infinito. Cada vez que miramos el electrón, éste está en alguna parte, pero tiene una amplitud de estar en diferentes lugares, de modo que hay una probabilidad de encontrarlo en distintos lugares. Esos lugares no pueden estar todos en el núcleo; supondremos que hay una distribución de la posición del orden a. Es decir, la distancia del electrón al núcleo es comúnmente alrededor de a. Determinaremos a minimizando la energía total del átomo.

El ancho de la distribución del momentum es de aproximadamente $\hbar/a$ debido a la relación de incertidumbre, de modo que si tratamos de medir en alguna forma el momentum del electrón, por ejemplo, mediante dispersión de Difracción de rayos X y mirando desde un dispersor móvil el efecto Doppler, no esperaríamos obtener cada vez cero —el electrón no está detenido—, pero los momenta deben ser del orden de $p \approx \hbar/a$. Entonces, la energía cinética es aproximadamente $\frac{1}{2}mv^2 = p^2/2m = \hbar^2/2ma^2$. (En cierto sentido, esto es una especie de análisis dimensional para encontrar en qué forma la energía cinética depende de la constante reducida de Planck, de m y del tamaño del átomo. No necesitamos confiar en nuestra respuesta a menos de factores como 2, π, etc. Ni siquiera hemos definido muy precisamente a.) Ahora bien, la energía potencial es menos e^2 dividido entre la distancia al centro, es decir $-e^2/a$, donde, recordemos, e^2 es el cuadrado de la carga del electrón, dividida entre $4\pi\epsilon_0$. El caso es ahora, que la energía potencial se reduce si a disminuye, pero mientras menor es a tanto mayor es el momentum necesario, debido al principio de incertidumbre, y por lo tanto mayor es la energía cinética, La energía total es:

$$E = \hbar^2/2ma^2 - e^2/a. \tag{38.10}$$

No sabemos cuánto es a, pero sabemos que el átomo se arreglará por sí mismo y hará cierto compromiso de modo que la energía sea tan pequeña como sea posible. Para minimizar E derivemos respecto de a; pongamos la derivada igual a cero y despejemos a. La derivada de E es:

$$dE/da = -\hbar^2/ma^3 + e^2/a^2, \tag{38.11}$$

y poniendo $dE/da = 0$, da a a el valor

$$\begin{aligned} a_0 = \hbar^2/me^2 &= 0.528\text{ángstrom}, \\ &= 0.528 \times 10^{-10}\text{metros}. \end{aligned} \tag{38.12}$$

Esta distancia particular se llama *radio de Borh*, y hemos aprendido así que las dimensiones atómicas son del orden de los ángstroms, lo cual es correcto. ¡Esto es algo estupendo: en realidad, es sorprendente, ya que hasta ahora no teníamos base para entender el tamaño de los átomos! Los átomos son algo completamente imposible desde el punto de vista clásico, ya que los electrones deberían caer en espiral hacia el núcleo.

Si ahora introducimos el valor (38.12) para a_0 en (38.10) para encontrar la energía, resulta:

$$E_0 = -e^2/2a_0 = -me^4/2\hbar^2 = -13.6 \text{ eV}. \tag{38.13}$$

¿Qué significa una energía negativa? Significa que el electrón tiene menor energía cuando está en el átomo que cuando está libre. Significa que está ligado. Significa que se requiere energía para expulsar el electrón, se requiere una energía del orden de 13.6 eV para ionizar un átomo de hidrógeno. No tenemos razón para pensar que no sea dos o tres veces esto —o la mitad de esto— o $(1/\pi)$ veces esto, ya que hemos usado un razonamiento tan poco cuidadoso. ¡Sin embargo, hemos trampeado, ya que hemos usado todas las constantes de tal manera de lograr que resulte el número exacto! Este número, 13.6 electronvoltios, se llama Rydbergs de energía, es la energía de ionización del hidrógeno.

De modo que ahora comprendemos por qué no caemos a través del piso. Cuando caminamos, nuestros zapatos, con sus masas de átomos, empujan contra el piso con *su* masa de átomos. Para apretar los átomos, los electrones deberían estar confinados en un espacio menor y, por el principio de incertidumbre, sus momenta deberían ser mayores en promedio, y esto significa mayor energía; la resistencia a la compresión atómica es un efecto cuántico y no un efecto clásico. Clásicamente, esperaríamos que si tuviéramos que juntar todos los electrones y protones más cerca unos de otros, la energía debería reducirse aún más, y el mejor arreglo de cargas positivas y negativas en la física clásica es de que todas estén una encima de la otra. Esto era en la física clásica y fue un rompecabezas debido a la existencia del átomo. Por supuesto, los primeros científicos inventaron algún camino para salir del problema —¡pero no se preocupen, nosotros tenemos *ahora* la salida correcta!— (Tal vez.)

A propósito, aunque no tenemos razón para entenderlo por el momento, en una situación donde hay muchos electrones, resulta que tratan de mantenerse alejados unos de otros. Si un electrón está ocupando un cierto espacio, otro no ocupa el mismo espacio. Más precisamente; hay dos maneras de girar, de modo que pueden ubicarse dos, uno arriba del otro, uno girando en un sentido y el otro en el otro sentido. Pero después de esto no podemos poner ninguno más. Tenemos que poner otros en otro lugar, y ésta es la razón verdadera de por qué la materia tiene resistencia. Si pudiéramos poner todos los electrones en el mismo lugar, se condensaría aún más de lo que está. Es el hecho de que los electrones no pueden estar uno encima del otro, que hace sólidas las masas y todo lo demás.

Es evidente que para entender las propiedades de la materia tendremos que usar la mecánica cuántica y no quedarnos satisfechos con la mecánica clásica.

38-5 Niveles de energía

Hemos hablado del átomo en su condición de más baja energía posible, pero resulta que el electrón puede hacer otras cosas. Puede moverse y agitarse de

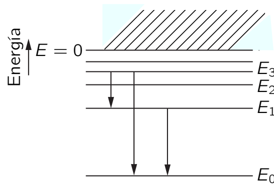


Figura 38-9. Diagrama energético para un átomo, mostrando varias transiciones posibles.

una manera más rápida y, por lo tanto, hay muchos movimientos diferentes y posibles para el átomo. De acuerdo con la mecánica cuántica, en una condición estacionaria puede haber solamente energías definidas para un átomo. Hacemos un diagrama (figura 38-9) en el cual dibujamos verticalmente la energía y trazamos una línea horizontal para cada valor permitido de la energía. Cuando el electrón está libre, es decir, cuando su energía es positiva, puede tener cualquier energía, se puede estar moviendo a cualquier velocidad. Pero las energías ligadas no son arbitrarias. El átomo debe tener uno u otro de un conjunto de valores permitidos, como los de la figura 38-9.

Llamemos E_0, E_1, E_2, E_3 a los valores permitidos de la energía. Si un átomo está inicialmente en uno de esos "estados excitados", E_1, E_2, *etc.*, no permanece en ese estado para siempre. Tarde o temprano caerá a un estado más bajo e irradiará energía en forma de luz. La frecuencia de la luz emitida está determinada por la conservación de la energía más el conocimiento cuántico de que la frecuencia de la luz está relacionada con la energía de la luz por la ecuación (38.1). Por lo tanto, la frecuencia de la luz liberada en una transición desde la energía E_3 a la energía E_1 (por ejemplo), es:

$$\omega_{31} = (E_3 - E_1)/\hbar. \tag{38.14}$$

Esto es entonces una frecuencia característica del átomo y define una línea de emisión espectral. Otra posible transición sería de E_3 a E_0. Tendría una frecuencia diferente,

$$\omega_{30} = (E_3 - E_0)/\hbar. \tag{38.15}$$

Otra posibilidad es que si el átomo fuera excitado al estado E_1, podría caer al estado fundamental E_0 emitiendo un fotón de frecuencia,

$$\omega_{10} = (E_1 - E_0)/\hbar. \tag{38.16}$$

La razón por la cual traemos a colación estas tres transiciones es para señalar una relación interesante. Es fácil ver a partir de (38.14), (38.15) y (38.16) que:

$$\omega_{30} = \omega_{31} + \omega_{10}. \tag{38.17}$$

En general, si encontramos dos líneas espectrales esperaremos hallar otra línea correspondiente a la suma de las frecuencias (o a la diferencia de las frecuencias) y que todas las líneas puedan ser entendidas encontrando una serie de niveles tales, que cada línea corresponde a la diferencia de energía de algún par de niveles. Esta coincidencia notable en las frecuencias espectrales fue notada antes de que la mecánica cuántica fuera descubierta, y se llama *principios de combinaciones de Ritz*. Éste es otra vez un misterio desde el punto de vista de la mecánica clásica. No insistamos en el punto de que la mecánica clásica es un fracaso en el dominio atómico, nos parece haberlo demostrado muy bien.

Hemos ya hablado sobre la mecánica cuántica como que es representada por amplitudes, las cuales se comportan como ondas, con ciertas frecuencias y números de onda. Observemos cómo llegamos desde el punto de vista de las amplitudes en que el átomo tiene estados energéticos definidos. Esto es algo que no podemos entender con lo que hemos dicho hasta aquí, pero todos estamos familiarizados con el hecho de que ondas encerradas tienen frecuencias definidas. Por ejemplo, si el sonido se limita a un tubo de órgano, o algo parecido, hay más de una forma en que el sonido pueda vibrar, pero para cada una de esas formas hay una frecuencia definida. Así, un objeto en el cual las ondas están encerradas tiene ciertas frecuencias de resonancia. Por lo tanto, es una propiedad de las ondas en un espacio confinado —una materia que discutiremos en detalle, con fórmulas, más tarde— que existan solamente a frecuencias definidas. Y ya que existe la relación general entre las frecuencias de la amplitud y la energía, no nos sorprende encontrar energías definidas asociadas con electrones ligados en los átomos.

38-6 Implicaciones filosóficas

Consideremos brevemente algunas implicaciones filosóficas de la mecánica cuántica. Como siempre, hay dos aspectos del problema: uno es la implicación filosófica para la física y el otro es la extrapolación de asuntos filosóficos a otros campos. Cuando las ideas filosóficas asociadas con la ciencia son arrastradas a otro campo, usualmente ellas son completamente distorsionadas. Por lo tanto, limitaremos nuestras consideraciones tanto como sea posible a la física misma.

En primer lugar, el aspecto más interesante es la idea del principio de incertidumbre, hacer una observación afecta el fenómeno. Siempre se ha sabido que hacer observaciones afecta un fenómeno, pero el punto es que el efecto no puede ser dejado a un lado, o minimizado o disminuido arbitrariamente mediante un nuevo arreglo del aparato. Cuando investigamos cierto fenómeno no nos queda más remedio que perturbarlo en cierta forma mínima, y *la perturbación es necesaria para la compatibilidad del punto de vista.* En la física anterior a la cuántica el observador fue algunas veces importante, pero solamente en un sentido más bien trivial. El problema ha sido establecido: si un árbol cae en un bosque y no hay nadie que lo escuche, ¿hace ruido? Un árbol *real* cayendo en un bosque *real*, hace, por supuesto, ruido aunque nadie esté ahí. Aunque no haya nadie presente para escucharlo, quedan otras huellas. El sonido agitará algunas hojas y si fuéramos lo suficientemente cuidadosos podríamos encontrar en alguna parte que cierta espina ha raspado contra una hoja y hecho una rayadura fina que no podría ser explicada, a menos que supongamos que la hoja estaba vibrando. De modo que en cierto sentido deberíamos admitir que hubo sonido. Podemos preguntarnos: ¿hubo una *sensación* de sonido? No, las sensaciones tienen que ver, presumiblemente, con la conciencia. Y si las hormigas son conscientes y si había hormigas en el bosque, o si los árboles tuvieran conciencia, no lo sabemos. Dejemos el problema en esa forma.

Otra cosa que la gente ha recalcado desde que la mecánica cuántica se ha desarrollado, es la idea de que no deberíamos hablar de las cosas que no podemos medir. (Realmente la teoría de la relatividad también decía esto.) A menos que una cosa pueda ser definida por la medida, no tiene cabida en una teoría. Y ya que un valor preciso del momentum de una partícula localizada no puede ser definido por medición, no tendrá, por lo tanto, cabida en la teoría. La idea de que esto es lo que sucedía con la teoría clásica *es una posición falsa.* Es un análisis sin cuidado de la situación. Que nosotros no podamos *medir* posición y momentum en forma precisa, no significa *a priori* que *no podamos* hablar de ello. Solamente significa que no *necesitamos* hablar de ello. La situación en las ciencias es ésta: un concepto o una idea que no pueden ser medidos o no pueden ser referidos directamente a un experimento pueden, o no pueden, ser útiles. No necesitan existir en una teoría. En otras palabras, supongamos que comparamos la teoría clásica del mundo con la teoría cuántica del mundo y supongamos que experimentalmente es verdad que podemos medir posición y momentum sólo en forma imprecisa. La cuestión es si los *conceptos* de posición exacta de una partícula y de momentum exacto de una partícula son válidos o no. La teoría clásica admite estos conceptos, la teoría cuántica no. Esto en sí no significa que la física clásica está equivocada. Cuando la nueva mecánica cuántica se descubrió, la gente clasicista —que incluía a todos, excepto a Heisenberg, Schrödinger y Born— dijo: "Miren, su teoría no es nada buena, porque ustedes no pueden contestar ciertas preguntas como: ¿cuál es la posición exacta de la partícula?, ¿por cuál hueco pasó?, y algunas otras más". La respuesta de Heisenberg fue: "Yo no necesito contestar tales preguntas, porque usted no puede experimentalmente formular tal pregunta". Es eso lo que no *tenemos* que hacer. Consideremos dos teorías (a) y (b), la (a) contiene una idea que no se puede verificar directamente, pero que se usa en el análisis, y la otra (b) no contiene la idea. Si están en desacuerdo en sus predicciones, no se podría pretender que (b) sea falsa, porque no puede explicar esta idea que está en (a), porque tal idea es una de las cosas que no se pueden verificar directamente. Es siempre bueno saber cuál de las ideas no puede ser verificada directamente, pero no es necesario eliminarlas todas.

No es verdad que nosotros podamos proseguir con la ciencia en forma completa usando sólo conceptos que están expuestos directamente al experimento.

En la mecánica cuántica misma existe una amplitud de función de onda, existe un potencial y existen muchas construcciones mentales que no pueden ser medidas directamente. La base de una ciencia está en su capacidad para *predecir*. Predecir significa decir qué sucederá en un experimento que no ha sido nunca hecho. ¿Cómo podemos hacer eso? Suponiendo que sabemos lo que hay, independientemente del experimento. Debemos extrapolar los experimentos a una región donde no han sido aún verificados. Si no hacemos eso, no tenemos predicción. De modo que era perfectamente razonable para los físicos clásicos proseguir con toda felicidad y suponer que la posición —la cual obviamente significa algo para una pelota de beisbol— significa algo también para un electrón. No era una estupidez. Era un procedimiento razonable. Hoy en día decimos que la ley de la relatividad se supone válida para todas las energías, pero algún día alguien puede llegar a decir lo estúpidos que fuimos. No sabemos en qué somos "estúpidos" hasta que no "vayamos hasta las últimas consecuencias", de modo que la idea global es ir hasta las últimas consecuencias. Y la única forma de encontrar que estábamos equivocados es encontrar *cuáles* son nuestras predicciones. Es absolutamente necesario construir.

Hemos expuesto ya algunas observaciones respecto a la incertidumbre de la mecánica cuántica. Es decir, que ahora somos incapaces de predecir lo que sucederá en física en una circunstancia física dada, que ha sido dispuesta lo más cuidadosamente posible. Si tomamos un átomo en un estado excitado, de modo que esté por emitir un fotón, no podemos decir *cuándo* lo emitirá. Tiene una cierta amplitud de emitir el fotón en cualquier instante y podemos predecir solamente una probabilidad de emisión, no podemos predecir exactamente el futuro. Esto ha dado origen a todo tipo de disparates y preguntas respecto al significado del libre albedrío, y a la idea de que el mundo es incierto.

Por supuesto, debemos recalcar que la mecánica clásica es también en cierto sentido indeterminada. Se piensa generalmente que tal incertidumbre, que no podamos predecir el futuro, es un importante asunto cuántico y se dice que esto explica el comportamiento de la mente, el sentimiento del libre albedrío, etc. Pero si el mundo *fuera* clásico —si las leyes de la mecánica fueran clásicas—, no es tan obvio que la mente no se sentiría más o menos igual. Es clásicamente verdadero que si conocemos la posición y la velocidad de cada partícula en el mundo, o en una caja con gas, podríamos predecir exactamente lo que sucedería. Y por lo tanto, el mundo clásico es determinista. Supongamos, sin embargo, que tenemos una precisión finita y no sabemos *exactamente* dónde está un átomo, digamos dentro de una parte en 1 000 millones. Entonces, a medida que avanza choca a otro átomo y, dado que no conocíamos la posición mejor que una parte en 1 000 millones, encontraremos un mayor error aún en la posición después del choque. Y esto, por supuesto, es amplificado en el choque siguiente, de modo que si comenzamos con sólo un ligero error éste aumenta rápidamente a una incertidumbre muy grande. Para dar un ejemplo: si el agua cae sobre una represa, salpica. Si nos paramos cerca, entonces de vez en cuando alguna gota caerá sobre nuestra nariz. Esto parece ocurrir completamente al azar; sin embargo, tal comportamiento podría ser predicho mediante leyes puramente clásicas. La posición exacta de todas las gotas depende de las ondulaciones precisas del agua antes de llegar a la represa. ¿Cómo? Las más pequeñas irregularidades son aumentadas en la caída, de modo que obtenemos una situación completamente al azar. Por supuesto, no podemos realmente predecir la posición de las gotas a menos que conozcamos el movimiento del agua en forma *absolutamente exacta*.

Hablando en forma más precisa, dada una precisión arbitraria, tan grande como se quiera, uno puede encontrar un tiempo suficientemente largo que no podemos hacer predicciones válidas para ese tiempo así de largo. Ahora bien, el asunto es que ese intervalo de tiempo no es muy grande. No se trata de que el tiempo sea millones de años si la precisión es de una parte en 1 000 millones. En efecto, el tiempo varía sólo logarítmicamente con el error y resulta que sólo en un intervalo de tiempo muy, pero muy pequeño, perdemos toda nuestra información.

Si la precisión se toma como una parte en miles de millones y miles de millones —no importa cuántos miles de millones queramos, siempre que nos detengamos en alguna parte—, podremos entonces encontrar un tiempo menor que el que se tomó para establecer la precisión ¡después del cual no podemos predecir más qué es lo que sucederá! Por lo tanto, no es honesto decir que de la libertad aparente y de la incertidumbre de la mente humana nos hubiéramos dado cuenta de que la física clásica "determinista" nunca podría esperar entenderlo, y dar la bienvenida a la mecánica cuántica como una liberación de un universo "completamente mecanicista". Porque ya en la mecánica clásica desde el punto de vista práctico hubo incertidumbre.

39

La teoría cinética de los gases

39-1 Propiedades de la materia

Con este capítulo comenzamos con un nuevo tema que nos ocupará algún tiempo. Es la primera parte del análisis de las propiedades de la materia bajo un punto de vista físico, en la que, reconociendo que la materia está hecha de una gran cantidad de átomos o partes elementales, que interactúan eléctricamente y obedecen las leyes de la mecánica, tratamos de entender por qué los diversos agregados de átomos se comportan en la forma que lo hacen.

Es evidente que éste es un tema difícil, y ponemos énfasis desde el comienzo en que es en realidad un tema *extremadamente* difícil, y que tenemos que tratarlo de manera diferente a como hemos tratado hasta ahora los otros temas. En el caso de la mecánica y en el caso de la luz pudimos comenzar con un enunciado preciso de algunas leyes, como las leyes de Newton, o la fórmula para el campo producido por una carga que acelera, a partir de las cuales pudimos comprender esencialmente un sinnúmero de fenómenos, y que dieron, desde ese momento, una base para nuestra comprensión de la mecánica y de la luz a partir de entonces. Esto es, más tarde podemos aprender más, que no será una física diferente, solamente aprenderemos mejores métodos de análisis matemático para tratar la situación.

No podemos usar este método en forma efectiva para el estudio de las propiedades de la materia. Podemos discutir la materia sólo en una forma más elemental, es un tema demasiado complicado para ser analizado directamente a partir de sus leyes básicas específicas, que no son otras que las leyes de la mecánica y de la electricidad. Pero éstas están bastante alejadas de las propiedades que deseamos estudiar, se requieren demasiados pasos para obtener a partir de las leyes de Newton las propiedades de la materia, y estos pasos son, en sí, bastante complicados. Comenzaremos ahora a dar algunos de estos pasos, pero mientras que muchos de nuestros análisis serán bastante exactos, se harán a la larga menos y menos exactos. Sólo tendremos una comprensión aproximada de las propiedades de la materia.

Una de las razones de que tengamos que realizar el análisis en forma tan imperfecta es que su matemática requiere una comprensión profunda de la teoría de probabilidades; no desearemos saber por dónde se mueve realmente cada átomo, sino, más bien, cuántos se mueven aquí y allá en promedio y cuáles son las probabilidades de diferentes efectos. Así, este tema comprende un conocimiento de la teoría de probabilidades y nuestra matemática todavía no está bien preparada y no deseamos forzarla demasiado.

En segundo lugar, y más importante desde el punto de vista físico, el comportamiento real de los átomos no es de acuerdo con la mecánica clásica, sino de acuerdo con la mecánica cuántica, y una comprensión correcta del tema no se puede alcanzar hasta que comprendamos la mecánica cuántica. Aquí, a diferencia del caso de las bolas de billar y de los automóviles, la diferencia entre las leyes de la mecánica clásica y las leyes de la mecánica cuántica es muy importante y muy significativa, de manera que muchas de las cosas que deduciremos por medio de la mecánica clásica serán fundamentalmente incorrectas. Por lo tanto, habrá ciertas cosas que tendrán que ser parcialmente desaprendidas; sin embargo, indicaremos en cada caso cuándo un resultado es incorrecto, de modo que sabremos exactamente dónde están los “límites”. Una de las razones para discutir la mecánica cuántica en los capítulos anteriores fue la de dar una idea de por qué, más o menos, la mecánica clásica es incorrecta en las diversas direcciones.

¿Por qué tratamos el tema ahora? ¿Por qué no esperamos medio año o un año, hasta que conozcamos mejor la matemática de las probabilidades, y que aprendamos un poco de mecánica cuántica, y entonces podamos hacerlo de una manera más fundamental? La respuesta es que es un tema difícil ¡y la mejor manera de aprenderlo es haciéndolo lentamente! Lo primero que hay que hacer es lograr alguna idea, más o menos, de lo que debiera suceder en diferentes circunstancias, y luego, cuando conozcamos mejor las leyes, las formularemos mejor.

Cualquiera que desee analizar las propiedades de la materia en un problema real, querría comenzar escribiendo las ecuaciones fundamentales y tratar luego de resolverlas matemáticamente. Aunque hay personas que tratan de usar este método, estas personas son las que fallan en este campo; el éxito real proviene de aquellos que comienzan desde un punto de vista *físico*, personas que tienen una idea aproximada de hacia dónde van y que luego comienzan haciendo el tipo correcto de aproximaciones, sabiendo qué es grande y qué es chico en una situación dada complicada. Estos problemas son tan complicados que es útil tener una comprensión siquiera elemental, aunque inexacta e incompleta, y así el tema será uno sobre el que volveremos una y otra vez, cada vez con más y más exactitud, a medida que avancemos en nuestro curso de física.

Otra razón para comenzar el tema precisamente ahora es que ya hemos usado muchas de estas ideas en otras áreas, por ejemplo en la química, y hasta hemos oído acerca de algunas de ellas en la segunda enseñanza. Es interesante conocer la base física de estas cosas.

Como un ejemplo interesante, todos sabemos que volúmenes iguales de gases a la misma presión y temperatura contienen el mismo número de moléculas. La ley de las proporciones múltiples, que cuando dos gases se combinan en una reacción química, los volúmenes necesarios están siempre en proporciones enteras simples, fue entendida en última instancia por Amedeo Avogadro como significando que volúmenes iguales contienen números iguales de átomos. Ahora bien, *¿por qué* contienen igual número de átomos? ¿Podemos deducir a partir de las leyes de Newton que el número de átomos debe ser igual? Nos preocuparemos de esta materia específica en este capítulo. En los capítulos siguientes discutiremos varios otros fenómenos que comprenden presiones, volúmenes, temperatura y calor.

Encontraremos también que el tema puede atacarse desde un punto de vista no atómico, y que existen muchas interrelaciones de las propiedades de las sustancias. Por ejemplo, si comprimimos alguna cosa, se calienta, si la calentamos, se dilata. Existe una relación entre estos dos hechos que se puede deducir independientemente del mecanismo involucrado. Este tema se llama *termodinámica*. La comprensión más profunda de la termodinámica proviene, por supuesto, de la comprensión del mecanismo real involucrado, y esto es lo que haremos: tomaremos el punto de vista atómico desde el comienzo y lo usaremos para comprender las diversas propiedades de la materia y las leyes de la termodinámica.

Discutamos, entonces, las propiedades de los gases desde el punto de vista de las leyes de la mecánica de Newton.

39-2 La presión de un gas

En primer lugar, sabemos que un gas ejerce una presión y debemos comprender claramente a qué se debe esto. Si nuestros oídos fueran unas pocas veces más sensibles, podríamos escuchar un zumbido perpetuo. La evolución no ha desarrollado el oído hasta ese punto, porque sería inútil si fuera tan sensible —oiríamos un jaraneo perpetuo—. La razón es que el tímpano está en contacto con el aire, y el aire consiste en muchas moléculas en movimiento perpetuo, y éstas golpean contra los tímpanos. Al golpear contra los tímpanos producen un tamborileo irregular —bum, bum, bum…— que no escuchamos porque los átomos son muy pequeños, y la sensibilidad del oído no es suficiente para notarlo. El resultado de este bombardeo perpetuo es desplazar la membrana, pero por supuesto existe un bombardeo perpetuo igual de átomos por el otro lado del tímpano; así la fuerza neta sobre él es cero. Si sacáramos el aire de un lado

o cambiáramos las cantidades relativas de aire en los dos lados, el tímpano desplazaría entonces hacia un lado u otro, porque el bombardeo por un lado sería mayor que por el otro. A veces sentimos este efecto molesto cuando subimos demasiado rápido en un ascensor o en un avión, especialmente cuando tenemos un resfrío grande (cuando estamos resfriados, la inflamación cierra el tubo que conecta el aire en el interior del tímpano con el aire exterior a través de la garganta, de modo que las dos presiones no pueden igualarse pronto).

Al considerar cómo analizar cuantitativamente la situación, imaginemos que tenemos en una caja un volumen de gas, en un extremo de la cual hay un pistón que se puede mover (figura 39-1). Nos gustaría encontrar qué fuerza resulta sobre el pistón del hecho que existen átomos en esta caja. El volumen de la caja es V, y a medida que los átomos se mueven en el interior de la caja con diferentes velocidades, golpean contra el pistón. Supongamos que no hay nada, un vacío, en el exterior del pistón. ¿Qué pasa? Si el pistón se dejara solo y nadie lo sostuviera, cada vez que fuera golpeado adquiriría un pequeño momentum y gradualmente sería empujado hacia fuera de la caja. De modo que para evitar que sea expulsado fuera de la caja debemos sostenerlo mediante una fuerza F. El problema es ¿cuánta fuerza? Una forma de expresar la fuerza es hablar de la fuerza por unidad de área; si A es el área del pistón, la fuerza sobre el pistón se escribirá como un número multiplicado por el área. Entonces, definimos la presión como igual a la fuerza que tenemos que aplicar al pistón, dividida entre el área del pistón:

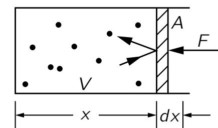


Figura 39-1. Átomos de un gas en una caja con un pistón sin roce.

$$P = F/A. \tag{39.1}$$

Para estar seguros de que entendemos la idea (tenemos que deducirla de todos modos para otros propósitos), el *trabajo* diferencial dW hecho sobre el gas al comprimirlo, moviendo el pistón en una cantidad diferencial $-dx$, sería el producto de la fuerza por la distancia que lo hemos comprimido, el cual sería de acuerdo con (39.1) el producto de la presión por el área multiplicada por la distancia, que es igual a menos el producto de la presión por la variación en el volumen:

$$dW = F(-dx) = -PA\,dx = -P\,dV. \tag{39.2}$$

(El producto del área A por la distancia dx es la variación de volumen.) El signo menos está ahí porque a medida que comprimimos, *disminuimos* el volumen; si pensamos en ello, podemos ver que, si se comprime un gas, se realiza un trabajo *sobre él*.

¿Cuánta fuerza tenemos que aplicar para equilibrar el golpeteo de las moléculas? El pistón recibe de cada colisión cierto momentum. Algún momentum por segundo se comunicará al pistón y éste comenzará a moverse. Debemos impartir en sentido contrario el mismo momentum por segundo, mediante nuestra fuerza, para evitar que se mueva. Por supuesto, la fuerza *es* el momentum por segundo que debemos comunicar hacia dentro. Hay otra forma de decirlo: si dejamos ir al pistón, adquirirá velocidad debido a los bombardeos; con cada colisión obtenemos un poco más de velocidad, por lo que la velocidad aumenta. La razón a la cual el pistón adquiere velocidad, o acelera, es proporcional a la fuerza sobre él. De modo que vemos que la fuerza que, como ya dijimos, es el producto de la presión por el área, es igual al momentum por segundo entregado al pistón por las moléculas que chocan.

Calcular el momentum por segundo es fácil, podemos hacerlo en dos partes: primero, encontramos el momentum entregado al pistón por un átomo particular durante una colisión con el pistón; a continuación, tenemos que multiplicar por el número de colisiones por segundo que los átomos realizan con la pared. La fuerza será el producto de esos dos factores. Ahora veamos cuáles son esos dos factores: en primer lugar supondremos que el pistón es un "reflector" perfecto para los átomos. Si no lo es, toda la teoría es errónea y el pistón comenzará a calentarse y las cosas cambiarán, pero eventualmente, cuando se haya establecido el equilibrio, el resultado neto es que las colisiones son, efectivamente, perfectamente elásticas. En promedio, cada partícula que llega, sale con la misma energía. De modo que

imaginaremos que el gas está en una condición estacionaria y que no perdemos energía en el pistón, porque el pistón está detenido. En esas circunstancias, si una partícula llega con una cierta velocidad sale con la misma velocidad, y diremos que con la misma masa.

Si $\boldsymbol{v}$ es la velocidad de un átomo y v_x es la componente x de $\boldsymbol{v}$, entonces mv_x es la componente x del momentum "hacia", pero también tenemos igual componente del momentum "desde" y así el momentum total entregado al pistón por la partícula en una colisión es $2mv_x$, ya que ésta se "refleja".

Ahora bien, necesitamos el número de colisiones hechas por los átomos en un segundo, o un cierto tiempo dt, entonces dividimos entre dt. ¿Cuántos átomos están golpeando? Supongamos que hay N átomos en el volumen V, o $n = N/V$ en cada unidad de volumen. Para encontrar cuántos átomos golpean el pistón notemos que, dado un cierto tiempo t, si una partícula tiene una cierta velocidad hacia el pistón, lo golpeará durante el tiempo t siempre que esté suficientemente cerca. Si está demasiado lejos, hará solamente parte del camino hacia el pistón en el tiempo t, pero no alcanza el pistón. Por lo tanto, está claro que solamente las moléculas que están dentro de una distancia v_xt del pistón golpearán el pistón en el tiempo t. Así, pues, el número de colisiones en un tiempo t es igual al número de átomos que están en una región dentro de una distancia v_xt y ya que el área del pistón es A, el *volumen* ocupado por los átomos que irán a golpear el pistón es v_xtA. Pero el *número* de átomos que irán a golpear el pistón es ese volumen multiplicado por el número de átomos por unidad de volumen, nv_xtA. Por supuesto, nosotros no queremos el número que golpea en un tiempo t, queremos el número que golpea por segundo, así que dividimos entre el tiempo t para obtener nv_xA. (Este tiempo t pudo haberse hecho muy corto; si deseamos ser más elegantes, lo llamaremos dt y entonces derivamos, pero es la misma cosa.)

De modo que encontramos que la fuerza es:

$$F = nv_xA \cdot 2mv_x. \tag{39.3}$$

¡Vean, la fuerza es proporcional al área si mantenemos fija la densidad de partículas a medida que variemos el área! La presión es, entonces:

$$P = 2nmv_x^2. \tag{39.4}$$

Ahora bien, notamos un pequeño problema con este análisis: primero, todas las moléculas no tienen la misma velocidad y no se mueven en la misma dirección. ¡De modo que todas las v_x^2 son diferentes! Así que lo que debemos hacer, por supuesto, es tomar un *promedio* de las v_x^2, ya que cada una hace su propia contribución. Lo que queremos es el cuadrado de las v_x promediado sobre todas las moléculas.

$$P = nm\langle v_x^2\rangle. \tag{39.5}$$

¿Olvidamos incluir el factor 2? No, de todos los átomos solamente la mitad se dirige hacia el pistón. La otra mitad está dirigida en otra dirección, por lo que el número de átomos por unidad de volumen *que está golpeando el pistón* es sólo $n/2$. El promedio de v_x^2 para v_x positivos es igual al promedio de las v_x^2 para *todo* v_x multiplicado por un medio.

Ahora bien, como los átomos rebotan por todas partes, está claro que no hay nada especial respecto a la "dirección x"; los átomos se pueden estar moviendo también hacia arriba y hacia abajo, hacia adelante y hacia atrás, hacia adentro y hacia afuera. Por lo tanto, será cierto que $\langle v_x^2\rangle$, el movimiento promedio de los átomos en una dirección y el promedio en las otras dos direcciones, serán todos iguales:

$$\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle. \tag{39.6}$$

Es solamente un asunto de matemática más bien artificiosa notar, por lo tanto, que son todos iguales a un tercio de su suma, que es, por supuesto, el cuadrado del módulo de la velocidad:

$$\langle v_x^2 \rangle = \tfrac{1}{3}\langle v_x^2 + v_y^2 + v_z^2 \rangle = \langle v^2 \rangle/3. \tag{39.7}$$

Esto tiene la ventaja de que no tenemos que preocuparnos de ninguna dirección particular, y así escribimos nuevamente nuestra fórmula para la presión en esta forma:

$$P = (\tfrac{2}{3})n\langle mv^2/2 \rangle. \tag{39.8}$$

La razón de que escribamos el último factor como $\langle mv^2/2 \rangle$ está en que esto es la *energía cinética* del movimiento del centro de masa de la molécula. Por lo tanto, encontramos que

$$PV = N(\tfrac{2}{3})\langle mv^2/2 \rangle. \tag{39.9}$$

Con esta ecuación podemos calcular cuál es la presión si conocemos las velocidades.

Como un ejemplo muy simple tomemos el gas helio o cualquier otro gas como vapor de mercurio, o vapor de potasio a temperatura suficientemente alta, o argón, en los cuales todas las moléculas son átomos simples, por lo cual podemos suponer que no hay movimiento interno en el átomo. Si tuviéramos una molécula compleja podría haber algún movimiento interno, vibraciones mutuas o algo. Supongamos que podemos despreciar eso; esto es realmente un asunto importante, al cual tendremos que volver, pero resulta ser correcto. Supongamos que el movimiento interno de los átomos se puede despreciar, y por lo tanto, para este propósito, la energía cinética del movimiento del centro de masa es toda la energía existente. De esta manera, para un gas monoatómico, la energía cinética es la energía total. En general llamaremos U a la energía total (algunas veces se llama energía total *interna*; podemos preguntarnos por qué, ya que no hay energía *externa* al gas), es decir, toda la energía de todas las moléculas del gas o del objeto o lo que sea.

Para un gas monoatómico supondremos que la energía U es igual al número de átomos multiplicado por el promedio de la energía cinética de cada uno, porque hemos despreciado cualquier posibilidad de excitación o movimiento interno en los átomos mismos. Entonces, en esas circunstancias, tendríamos

$$PV = \tfrac{2}{3}U. \tag{39.10}$$

A propósito, podemos detenernos aquí y encontrar la respuesta a la siguiente pregunta: supongamos que tomamos una lata con gas y comprimimos lentamente el gas, ¿cuánta presión necesitamos para reducir el volumen? Es fácil calcularlo ya que la presión es $\frac{2}{3}$ de la energía, dividido entre V. A medida que lo comprimimos estamos realizando un trabajo sobre el gas, y por lo tanto aumentamos la energía U. De modo que tendremos algún tipo de ecuación diferencial: si comenzamos en una circunstancia dada con una cierta energía y volumen, conocemos la presión. Ahora comenzamos a comprimir, pero en el momento que lo hacemos, la energía U aumenta y el volumen V disminuye, de modo que la presión aumenta.

Así, tenemos que resolver una ecuación diferencial y la resolveremos dentro de un momento. Primeramente debemos hacer, sin embargo, énfasis en que, a medida que comprimimos este gas, estamos suponiendo que todo el trabajo va a aumentar la energía de los átomos internos. Podemos preguntarnos: “¿Acaso no es eso necesario? ¿Dónde más pudo ir?” Resulta que *puede* ir a otra parte, existe lo que llamamos “pérdida de calor” a través de las paredes: los átomos calientes (es decir, de movimiento rápido) que bombardean las paredes, calientan las paredes, y la energía se va. Supondremos por ahora que éste no es el caso.

Para mayor generalidad, aunque todavía estamos haciendo algunas consideraciones muy especiales respecto a nuestro gas, escribiremos no $PV = \frac{2}{3}U$, sino

$$PV = (\gamma - 1)U. \tag{39.11}$$

Está escrito $(\gamma - 1)$ multiplicado por U por razones convencionales, porque trataremos más tarde algunos otros casos donde el número frente a U no será

$\frac{2}{3}$ sino que será un número diferente. Así, para hacer la cosa en general, lo llamaremos $\gamma - 1$, porque la gente lo ha estado llamando así por casi 100 años. Este γ es entonces $\frac{5}{3}$ para un gas monoatómico como el helio, porque $\frac{5}{3} - 1$ es $\frac{2}{3}$.

Ya hemos hecho notar que cuando comprimimos un gas, el trabajo realizado es $-PdV$. Una compresión en la cual no se ha agregado ni sacado energía térmica, se llama compresión *adiabática*, del griego *a* (no) + *día* (a través) + *bainein* (ir). (La palabra adiabática se usa en física de varias maneras y algunas veces es difícil ver qué tienen de común entre ellas.) Esto es, para una compresión adiabática, todo el trabajo realizado irá a variar la energía interna. Ésta es la clave —que no hay otras pérdidas de energía— pues entonces tenemos $PdV = -dU$. Pero como $U = PV/(\gamma - 1)$ podemos escribir

$$dU = (P\,dV + V\,dP)/(\gamma - 1). \tag{39.12}$$

De modo que tenemos $PdV = -(PdV + VdP)/(\gamma - 1)$, o reagrupando los términos $\gamma PdV = -VdP$, o sea

$$(\gamma\,dV/V) + (dP/P) = 0. \tag{39.13}$$

Afortunadamente, suponiendo que γ es constante, como lo es para un gas monoatómico, podemos integrar esto: se obtiene $\gamma \ln V + \ln P = \ln C$, donde $\ln C$ es la constante de integración. Si tomamos exponenciales de ambos lados, obtenemos a ley

$$PV^\gamma = C \text{ (una constante)}. \tag{39.14}$$

¡En otras palabras, en condiciones adiabáticas donde la temperatura aumenta a medida que comprimimos porque no hay pérdidas calóricas, la presión multiplicada por el volumen a la potencia $\frac{5}{3}$ es una constante para un gas monoatómico! Aunque lo dedujimos teóricamente, ésta *es*, en realidad, la forma en que los gases monoatómicos se comportan experimentalmente.

39-3 Compresibilidad de la radiación

Podemos dar otro ejemplo de la teoría cinética de un gas, uno que no se usa mucho en química, pero sí en astronomía. Tenemos un gran número de fotones en una caja en la cual la temperatura es muy alta. (Por supuesto, la caja es el gas en una estrella muy caliente. El Sol no es lo suficientemente caliente; hay todavía muchos átomos, pero a temperaturas aún mayores, en ciertas estrellas muy calientes podemos omitir los átomos y suponer que los únicos objetos que tenemos en la caja son fotones.) Ahora bien, un fotón tiene cierto momentum $\boldsymbol{p}$. (Siempre encontramos que estamos en un terrible dilema cuando hacemos teoría cinética: p es la presión, pero p es el momentum; v es el volumen, pero v es la velocidad; T es la temperatura, pero T es la energía cinética, o el tiempo, o el torque; ¡uno debe mantenerse alerta respecto a ello!) Este $\boldsymbol{p}$ es el momentum, es un vector. Realizando el mismo análisis anterior, es la componente x del vector $\boldsymbol{p}$ la que genera el “golpe” y el doble de la componente x del vector $\boldsymbol{p}$ es el momentum entregado en el golpe. Así, $2p_x$ remplaza a $2mv_x$, y, en el cálculo del número de colisiones, v_x es todavía v_x, de modo que cuando hacemos todo el camino, encontramos que la presión en la ecuación (39.4) es, en cambio,

$$P = 2np_xv_x. \tag{39.15}$$

En promedio, entonces, esto llega a ser n veces el promedio de p_xv_x (el mismo factor 2) y finalmente, introduciendo las otras dos direcciones, encontramos

$$PV = N\langle \boldsymbol{p} \cdot \boldsymbol{v} \rangle/3. \tag{39.16}$$

Esto concuerda con la fórmula (39.9) porque el momentum es $m\boldsymbol{v}$; esto es un poco más general, eso es todo. La presión multiplicada por el volumen es el promedio del producto entre el número total de átomos y $\frac{1}{3}(\boldsymbol{p} \cdot \boldsymbol{v})$.

Ahora bien, ¿qué es $\boldsymbol{p} \cdot \boldsymbol{v}$ para los fotones? El momentum y la velocidad tienen la misma dirección y la velocidad es la de la luz, de modo que aquélla es el momentum de cada uno de los objetos, multiplicado por la velocidad de la luz. El momentum, multiplicado por la velocidad de la luz de cada fotón, es su energía: $E = pc$, de modo que esos términos son las *energías* de cada uno de los fotones y, por supuesto, deberíamos tomar una energía promedio, multiplicada por el número de fotones. Así tenemos $\frac{1}{3}$ de la energía dentro del gas:

$$PV = U/3 \text{ (gas de fotones).} \tag{39.17}$$

Entonces para fotones $(\gamma - 1)$ en (39.11) es $\frac{1}{3}$, ya que tenemos $\frac{1}{3}$ al frente, o sea $\gamma = \frac{4}{3}$, y hemos descubierto que la radiación en una caja obedece la ley

$$PV^{4/3} = C. \tag{39.18}$$

¡De modo que conocemos la compresibilidad de la radiación! Esto es lo que se usa en un análisis de la contribución de la presión de radiación en una estrella, así es cómo la calculamos y cómo cambia cuando la comprimimos. ¡Qué cosas maravillosas están ya dentro de nuestro poder!

39-4 Temperatura y energía cinética

Hasta aquí no nos hemos preocupado de la *temprratura*; hemos estado evitando a propósito la temperatura. A medida que comprimimos un gas sabemos que la energía de las moléculas aumenta, y acostumbramos a decir que el gas se calienta; nos gustaría entender qué tiene que ver esto con la temperatura. ¿Qué estamos haciendo, si tratamos de hacer el experimento, no adiabáticamente, sino a lo que llamamos *temperatura constante*? Sabemos que si tomamos dos cajas de gas y las ubicamos una al lado de la otra por un tiempo suficientemente largo, aunque en un comienzo estuvieran a lo que llamamos temperaturas diferentes, terminarán por llegar a la misma temperatura. Ahora bien, ¿qué significa eso? ¡Eso significa que llegaron a una condición igual a la que llegarían si se los dejara solos un tiempo suficientemente largo! Lo que queremos decir por igual temperatura es sólo eso, la condición final cuando las cosas se han asentado interactuando entre sí durante un tiempo suficientemente largo.

Consideremos ahora lo que sucede si tenemos dos gases en depósitos separados por un pistón móvil, como en la figura 39-2 (sólo por simplicidad tomaremos dos gases monoatómicos como helio y neón). En el depósito (1) los átomos tienen masa m_1, velocidad v_1 y hay n_1 por unidad de volumen, y en el otro depósito los átomos tienen masa m_2, velocidad v_2 y hay n_2 átomos por unidad de volumen. ¿Cuáles son las condiciones de equilibrio?

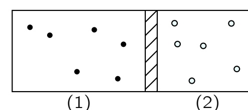


Figura 39-2. Átomos de dos gases monoatómicos diferentes están separados por un pistón móvil.

Es evidente que el bombardeo desde el lado izquierdo debe ser tal, que mueve el pistón hacia la derecha y comprime el otro gas hasta que su presión adquiere cierto valor y el objeto se moverá hacia adelante y atrás, y gradualmente llegará a detenerse en un lugar donde las presiones a ambos lados sean iguales. Así podemos hacer que las presiones sean iguales; esto sólo significa que las energías internas por unidad de volumen son iguales o que los números n multiplicados por el promedio de las energías cinéticas a cada lado son iguales. Lo que debemos tratar de probar finalmente es que los *números mismos* son iguales. Hasta aquí todo lo que sabemos es que los productos de los números y las energías cinéticas son iguales:

$$n_1 \langle m_1 v_1^2/2 \rangle = n_2 \langle m_2 v_2^2/2 \rangle,$$

según (39.8), porque las presiones son iguales. Debemos darnos cuenta que ésta no es la única condición a la larga, sino que algo más debe suceder más lentamente, a medida que el verdadero equilibrio completo correspondiente a temperaturas iguales se establece.

Para comprender el concepto, supongamos que la presión sobre el lado izquierdo se desarrollará debido a una muy alta densidad, pero baja velocidad.

Teniendo una gran n y una pequeña v podemos obtener la misma presión que si tuviéramos una pequeña n y una gran v. Los átomos pueden estar moviéndose lentamente ya que están agrupados casi sólidamente, o bien puede haber menos, pero golpeando más fuerte. ¿Podrá permanecer esto así para siempre? A primera vista podemos pensar eso, pero pensándolo otra vez encontramos haber olvidado un punto importante. Esto es, que el pistón intermedio no recibe una presión constante; vibra justamente como el tímpano, del cual hemos hablado primero, porque los golpeteos no son absolutamente uniformes. No hay una perpetua y constante presión, sino un tamborileo —la presión varía y por lo tanto el pistón se sacude—. Supongamos que los átomos del lado derecho no se están agitando mucho, pero los de la izquierda son pocos, muy alejados entre sí y muy rápidos. El pistón recibirá de vez en cuando un gran impulso desde la izquierda y será impulsado contra los átomos lentos de la derecha, dándoles más velocidad. (A medida que cada átomo choca con el pistón, éste o gana o pierde energía, dependiendo de si el pistón se mueve en uno u otro sentido cuando el átomo lo golpea.) De modo que como resultado de las colisiones el pistón se sacude, se sacude, se sacude, y esto agita al otro gas: le da energía a los otros átomos y ellos establecen movimientos más rápidos hasta que equilibran la agitación que el pistón les está dando. El sistema llega a algún equilibrio donde el pistón se está moviendo con una velocidad media cuadrática tal, que saca energía de los átomos en aproximadamente la misma proporción que les devuelve energía. De manera que el pistón adquiere una cierta irregularidad media en la velocidad y nuestro problema es encontrarla. Cuando la encontramos podemos resolver mejor nuestro problema, porque los gases ajustarán sus velocidades hasta que la rapidez a la cual estén tratando de entregar energía uno al otro, a través del pistón, llegue a ser igual.

Es bien difícil imaginarse los detalles del pistón en esta circunstancia particular; aunque es simple entenderlo idealmente resulta que es un poco más difícil analizarlo. Antes de hacer el análisis, estudiemos otro problema en el cual tenemos una caja de gas, pero ahora con dos tipos diferentes de moléculas en ella, que tienen masas m_1 y m_2, velocidades v_1 y v_2, etc.; ahora hay una relación mucho más íntima. Si todas las moléculas núm. 2 están detenidas, tal condición no va a durar, porque son golpeadas por las moléculas núm. 1 y así adquieren velocidad. Si todas fueran mucho más rápidas que las moléculas núm. 1, puede ser que eso tampoco dure —devolverán la energía a las moléculas núm. 1. De modo que cuando ambos gases están en la misma caja, el problema es encontrar la regla que determina las velocidades relativas de los dos.

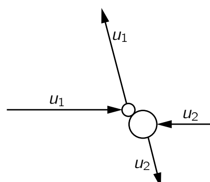


Figura 39-3. Una colisión entre moléculas desiguales, vista desde el sistema CM. $u_1 = |\boldsymbol{v}_1 - \boldsymbol{v}_{\text{CM}}|$, $u_2 = |\boldsymbol{v}_2 - \boldsymbol{v}_{\text{CM}}|$.

Éste es todavía un problema muy difícil, pero lo solucionaremos en la forma siguiente. Primero, consideraremos el subproblema siguiente (otra vez se tiene uno de aquellos casos donde —cualquiera que sea la deducción— al final el resultado es bastante simple de recordar, pero su deducción es bastante ingeniosa). Supongan que tenemos dos moléculas, de diferente masa, chocando y que la colisión se observa respecto al sistema del centro de masa (CM). Para eliminar una complicación, examinamos la colisión en el CM. Como sabemos por las leyes de colisiones, debido a la conservación del momentum y de la energía, después que las moléculas chocan el único modo en que pueden moverse es aquel en que cada una mantiene su propia velocidad original —sólo cambian su *dirección*—. De modo que tenemos una colisión promedio que se parece a la de la figura 39-3. Supongan, por un momento, que miramos todas las colisiones con el CM en reposo. Imagínense que todas se están moviendo inicialmente en forma horizontal. Por supuesto, después de la primera colisión algunas se están moviendo oblicuamente. Es decir, si todas ellas fueran horizontalmente, por lo menos algunas de ellas más tarde se moverían verticalmente. Ahora bien, en alguna otra colisión vendrían desde otra dirección y entonces se desviarían aún a otro ángulo. Así, aunque estuvieran en un comienzo completamente organizadas, se dispersarían en todos los ángulos, y luego las dispersadas se dispersarían algo más y algo más y algo más. Finalmente, ¿cuál será la distribución? *Respuesta: será igualmente probable encontrar un par cualquiera moviéndose en una dirección cualquiera en el espacio.* Después de esto, las colisiones posteriores no podrían cambiar la distribución.

Tienen la misma probabilidad para ir en todas direcciones, pero ¿cómo decimos eso? Por supuesto *no* hay ninguna probabilidad de que vayan en una dirección específica cualquiera, ya que una dirección específica es demasiado exacta, de modo que tenemos que hablar de una unidad de “algo”. La idea es que cualquier área sobre una esfera centrada en el punto de colisión tendrá tantas moléculas atravesándola como cualquier otra área igual sobre la esfera. De manera que el resultado de las colisiones será distribuir las direcciones de manera que áreas iguales sobre la esfera tendrán probabilidades iguales.

A propósito, si sólo queremos discutir la dirección original y alguna otra dirección en un ángulo θ respecto a ella, es una propiedad interesante que el área diferencial de una esfera de radio unitario es $\operatorname{sen}\theta\,d\theta$ multiplicado por 2π (véase figura 32-1), y que $\operatorname{sen}\theta\,d\theta$ es lo mismo que el diferencial de $-\cos\theta$. Así, lo que esto significa es que es igualmente probable que el coseno del ángulo θ entre dos direcciones cualesquiera tenga cualquier valor entre -1 y $+1$.

A continuación, tenemos que preocuparnos del caso real, donde no tenemos la colisión en el sistema CM, sino que tenemos dos átomos que se acercan con velocidades vectoriales $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ ¿Qué sucede ahora? Podemos analizar esta colisión con las velocidades vectoriales $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ en la siguiente forma: primero diremos que hay un cierto CM; la velocidad del CM está dada por la velocidad “promedio”, con los pesos proporcionales a las masas, de modo que la velocidad del CM es $\boldsymbol{v}_{CM} = (m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2/(m_1 + m_2)$. Si observamos esta colisión en el sistema CM, vemos una colisión justamente como en la figura 39-3, con una cierta velocidad relativa $\boldsymbol{w}$. La velocidad relativa es justamente $\boldsymbol{v}_1 - \boldsymbol{v}_2$. Ahora la idea es, primero, que todo el *CM* se está moviendo y en el CM hay una velocidad relativa $\boldsymbol{w}$, y las moléculas chocan y salen en una nueva dirección. Todo esto sucede mientras el CM continúa moviéndose sin ningún cambio.

Ahora bien, ¿cuál es la distribución que resulta de esto? De nuestros razonamientos anteriores concluimos esto: que en el equilibrio *todas las direcciones para* $\boldsymbol{w}$ *son igualmente probables respecto a la dirección del movimiento del* CM.* No habrá finalmente correlación particular alguna entre la dirección del movimiento de la velocidad relativa y la del movimiento del CM. Por supuesto, si la hubiera, las colisiones la esparcirían de modo que todo se derramaría en derredor. Así, el coseno del ángulo entre $\boldsymbol{w}$ y $\boldsymbol{v}_{\text{CM}}$ es cero en promedio. Esto es,

$$\langle \boldsymbol{w} \cdot \boldsymbol{v}_{\text{CM}} \rangle = 0. \tag{39.19}$$

Pero $\boldsymbol{w} \cdot \boldsymbol{v}_{\text{CM}}$ se puede también expresar en función de $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$:

$$\begin{aligned}\boldsymbol{w} \cdot \boldsymbol{v}_{\text{CM}} &= \frac{(\boldsymbol{v}_1 - \boldsymbol{v}_2)\cdot(m_1\boldsymbol{v}_1 + m_2\boldsymbol{v}_2)}{m_1 + m_2}\\ &= \frac{(m_1v_1^2 - m_2v_2^2) + (m_2 - m_1)(\boldsymbol{v}_1 \cdot \boldsymbol{v}_2)}{m_1 + m_2}.\end{aligned} \tag{39.20}$$

Examinemos primero el producto $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2$; ¿cuál es el promedio de $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2$? Esto es: ¿cuál es el promedio de las componentes de la velocidad de una molécula en la dirección de otra? Seguramente hay tanta probabilidad de encontrar cualquier molécula dada moviéndose en una forma u otra. *El promedio de la velocidad* $\boldsymbol{v}_2$ *en cualquier dirección es cero.* Entonces en la dirección de $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ tiene promedio cero. De modo que ¡el promedio de $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2$ es cero! Por lo tanto, concluimos que el promedio de $m_1v_1^2$ debe ser igual al promedio de $m_2v_2^2$. Es decir, *las energías cinéticas promedio de ambas deben ser iguales:*

$$\langle \tfrac{1}{2}m_1v_1^2 \rangle = \langle \tfrac{1}{2}m_2v_2^2 \rangle. \tag{39.21}$$

* Este razonamiento, que fue el usado por Maxwell, encierra algunas sutilezas. Aunque la conclusión es correcta, el resultado *no* se deduce puramente de las consideraciones de simetría que hemos usado antes, ya que al pasar a un sistema de referencia que se mueve a través del gas, podemos encontrar una distribución distorsionada de velocidades. No hemos encontrado una demostración simple de este resultado.

Si tenemos dos tipos de átomos en un gas, se puede demostrar, y supondremos que lo hemos demostrado, que el promedio de la energía cinética de uno es igual al promedio de la energía cinética del otro, cuando ambos estén en el mismo gas, en la misma caja, en equilibrio. Esto significa que el más pesado se moverá más lento que el más liviano; esto es fácilmente demostrable mediante experimentación con "átomos" de diferentes masas en un canal de aire.

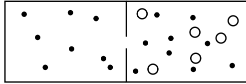


Figura 39-4. Dos gases en una caja con una membrana semipermeable.

Nos gustaría ahora avanzar un paso más y decir que, si tenemos dos gases diferentes, *separados* en una caja, tendrán también iguales energías cinéticas promedio cuando llegan finalmente al equilibrio, aunque no estén en la misma caja. Podemos hacer el razonamiento en muchas formas. Una forma es aducir que si tenemos una separación fija con un pequeño agujero en ella (figura 39-4) de modo que un gas pueda pasar a través del agujero, mientras que el otro no, porque las moléculas son demasiado grandes y éstas alcanzaron el equilibrio, entonces sabemos que en una parte donde están mezcladas tienen la misma energía cinética promedio, pero algunas atraviesan el agujero sin pérdida de energía cinética, de modo que la energía cinética promedio en el gas puro y en la mezcla debe ser la misma. Esto no es muy satisfactorio, porque puede no haber agujeros, para este tipo de moléculas, que separen un tipo de otro.

Volvamos ahora al problema del pistón. Demos un argumento que muestre que la energía cinética de este pistón debe ser también $\frac{1}{2}m_2v_2^2$. En realidad, ésta sería la energía cinética debida al movimiento puramente horizontal del pistón, de modo que, olvidando su movimiento hacia arriba y hacia abajo, debe ser igual a $\frac{1}{2}m_2v_{2_x}^2$. Análogamente, del equilibrio al otro lado, podemos deducir que la energía cinética del pistón es $\frac{1}{2}m_1v_{1_x}^2$. A pesar de que éste no está en el medio del gas sino a un lado del gas, aún podemos hacer el razonamiento, aunque sea algo más difícil, de que la energía cinética promedio del pistón y de las moléculas del gas son iguales como resultado de todas las colisiones.

Si esto aún no nos satisface, podemos hacer un ejemplo artificial mediante el cual el equilibrio es generado por un objeto que puede golpear sobre todos los lados. Supongamos tener una barra corta con una bola en cada extremo que atraviesa el pistón por una articulación universal, que desliza sin roce. Cada bola es redonda como una de las moléculas y puede ser golpeada por todos lados. El objeto completo tiene una cierta masa total m. Tenemos ahora las moléculas del gas con masas m_1 y m_2 como antes. El resultado de las colisiones, por el análisis que hemos hecho antes, es que la energía cinética de m, debido a las colisiones con las moléculas por un lado, debe ser $\frac{1}{2}m_1v_1^2$ en promedio. Análogamente, debido a las colisiones con moléculas en el otro lado, tiene que ser $\frac{1}{2}m_2v_2^2$ en promedio. En consecuencia, ambos lados tienen que poseer la *misma* energía cinética cuando están en equilibrio térmico. Así, aunque solamente lo hemos demostrado para una mezcla de gases, se puede extender fácilmente al caso en que hay dos gases diferentes, separados, a la misma temperatura.

Así, *cuando tenemos dos gases a la misma temperatura, la energía cinética media de los movimientos del CM son iguales.*

La energía cinética media molecular es una propiedad sólo de la "temperatura". Siendo una propiedad de la "temperatura" *y no del ley de Ohm*, podemos usarla como una *definición* de temperatura. La energía cinética media de una molécula es así alguna función de la temperatura. Pero ¿quién nos va a decir qué escala usar para la temperatura? Podemos *definir* arbitrariamente la escala de temperatura, tal que la energía media sea linealmente proporcional a la temperatura. La mejor manera de hacerlo sería llamar a la energía media misma "la temperatura". Ésta sería la función más simple posible. Desgraciadamente, la escala de temperatura se ha elegido en forma diferente; así que, en vez de llamarla directamente temperatura, usamos un factor de conversión constante entre la energía de una molécula y un grado de temperatura absoluta, llamado grado Kelvin. La constante de proporcionalidad es $k = 1.38 \times 10^{-23}$ joules por cada grado Kelvin.* Así, si T es la temperatura absoluta nuestra definición dice que la energía cinética

* La escala centígrada es precisamente esta escala Kelvin con el cero elegido a 273.15 K, así $T = 273.15 +$ temperatura centígrada.

molecular media es $\frac{3}{2}kT$. (El $\frac{3}{2}$ se ha puesto por conveniencia, para librarnos de él en alguna otra parte.)

Subrayamos que la energía cinética asociada con la componente del movimiento en cualquier dirección particular es solamente $\frac{1}{2}kT$. Las tres direcciones independientes implicadas la hacen $\frac{3}{2}kT$.

39-5 La ley de los gases ideales

Ahora, podemos poner nuestra definición de temperatura en la ecuación (39.9) y encontrar así la ley para la presión de gases, en función de la temperatura: ésta es, que la presión multiplicada por el volumen es igual al número total de átomos, multiplicado por la constante universal k, multiplicado por la temperatura.

$$PV = NkT. \tag{39.22}$$

Además, a la misma temperatura, presión y volumen se determina el *número de átomos*; ¡éste es también una constante universal! Así, volúmenes iguales de gases diferentes, a la misma presión y temperatura, contienen el mismo número de moléculas debido a las leyes de Newton. ¡Ésta es una conclusión asombrosa!

En la práctica, al tratar con moléculas, debido a que los números son tan grandes, los químicos han elegido artificialmente un número específico, un número muy grande, y lo llamaron de alguna manera. Tienen un número que llaman *mol*. Un mol es simplemente un número práctico. Por qué no eligieron 10^{24} objetos, para que resultara redondo, es una cuestión histórica. Sucedió que eligieron, para el número conveniente de objetos a tomar como referencia, $N_0 = 6.02 \times 10^{23}$ objetos, y eso se llama un mol de objetos. Así, en vez de medir el número de moléculas en unidades, miden en términos de números de moles.* En términos de N_0 podemos escribir el número de moles, multiplicado por el número de átomos en un mol, multiplicado por kT, y si deseamos, podemos tomar el número de átomos en un mol, multiplicado por k, que es un equivalente en moles de k y llamarlo de alguna manera, y lo hacemos —lo llamamos R—. El equivalente en moles de k es $8,317$ joules : $R = N_0 k = 8.317 j \cdot \mathrm{mol}^{-1} \cdot \mathrm{K}^{-1}$. Así encontramos también la ley de los gases escrita como el número de moles (también llamado N) multiplicado por RT, o el número de átomos multiplicado por kT:

$$PV = NRT. \tag{39.23}$$

Es la misma cosa, sólo una escala diferente para la medida de los números. ¡Usamos 1 como unidad, y los químicos usan 6×10^{23} como unidad!

Hacemos ahora una observación más sobre nuestra ley de los gases; tiene que ver con la ley para objetos que no son moléculas monoatómicas. Hemos tratado sólo el movimiento del CM de los átomos de un gas monoatómico. ¿Qué sucederá si hay fuerzas presentes? En primer lugar, consideremos el caso en que el pistón se sujeta por un resorte horizontal y que existen fuerzas sobre él. El intercambio del movimiento de agitación entre los átomos y el pistón en cualquier momento no depende de dónde está el pistón en ese momento, por supuesto. Las condiciones de equilibrio son las mismas. No importa dónde se encuentre el pistón, la velocidad del movimiento debe ser tal que pase energía a las moléculas justamente de la manera correcta. Así, no tiene ninguna importancia el resorte. La *velocidad* a la cual debe moverse el pistón es, en promedio, la misma. Así, nuestro teorema de que el valor medio de la energía cinética en una dirección es $\frac{1}{2}kT$, *es válido haya o no fuerzas presentes.*

Consideremos, por ejemplo, una molécula diatómica, compuesta de átomos m_A y m_B. Lo que hemos demostrado es que el movimiento del CM de la parte A y el de la parte B es tal que $\langle \frac{1}{2} m_A v_A^2 \rangle = \langle \frac{1}{2} m_B v_B^2 \rangle = \frac{3}{2}kT$. ¿Cómo puede ser esto si están unidas? A pesar de que están unidas, cuando giran en torno

* Lo que los químicos llaman pesos moleculares son las masas en gramos de un mol de una molécula determinada. El mol se define de manera tal que la masa de un mol de átomos de carbono del isótopo 12 (es decir, que tiene 6 protones y 6 neutrones en el núcleo) sea exactamente 12 gramos.

a sí mismas y una en torno a la otra, cuando algo las golpea, intercambiando energía con ellas, *lo único que cuenta es con qué velocidad se están moviendo*. Eso sólo determina lo rápido que intercambian energía en las colisiones. En el instante particular, la fuerza no es algo esencial. Por lo tanto, el mismo principio es correcto aun cuando haya fuerzas.

Demostremos, finalmente, que la ley de los gases es compatible también en la emisión del movimiento interno. No incluimos antes realmente los movimientos internos; tratamos sólo un gas monoatómico. Pero demostraremos ahora que un objeto completo, considerado como un único cuerpo de masa M, tiene una velocidad del CM tal que

$$\left\langle \tfrac{1}{2} M v_{\mathrm{CM}}^2 \right\rangle = \tfrac{3}{2} kT. \tag{39.24}$$

¡En otras palabras, podemos considerar las partes separadamente o el objeto entero! Veamos la razón de esto: la masa de la molécula diatómica es $M = m_A + m_B$ y la velocidad del centro de masas es igual a $\boldsymbol{v}_{\mathrm{CM}} = (m_A \boldsymbol{v}_A + m_B \boldsymbol{v}_B)/M$. Ahora necesitamos $\langle v_{\mathrm{CM}}^2 \rangle$. Si elevamos $\boldsymbol{v}_{\mathrm{CM}}$ al cuadrado, obtenemos

$$v_{\mathrm{CM}}^2 = \frac{m_A^2 v_A^2 + 2 m_A m_B \boldsymbol{v}_A \cdot \boldsymbol{v}_B + m_B^2 v_B^2}{M^2}.$$

Ahora multiplicamos por $\frac{1}{2}M$ y tomamos el promedio, y obtenemos así

$$\begin{aligned} \left\langle \tfrac{1}{2} M v_{\mathrm{CM}}^2 \right\rangle &= \frac{m_A \frac{3}{2} kT + m_A m_B \langle \boldsymbol{v}_A \cdot \boldsymbol{v}_B \rangle + m_B \frac{3}{2} kT}{M} \\ &= \tfrac{3}{2} kT + \frac{m_A m_B \langle \boldsymbol{v}_A \cdot \boldsymbol{v}_B \rangle}{M}. \end{aligned}$$

[Hemos usado el hecho de que $(m_A + m_B)/M = 1$.] Ahora bien, ¿qué es $\langle \boldsymbol{v}_A \cdot \boldsymbol{v}_B \rangle$? (¡Mejor que sea cero!) Para averiguarlo, usemos nuestra hipótesis de que la velocidad relativa, $\boldsymbol{w} = \boldsymbol{v}_A - \boldsymbol{v}_B$, probablemente no está dirigida más en una dirección que en otra —dicho de otro modo, que su componente promedio en cualquier dirección es cero—. Así, pues, suponemos que

$$\langle \boldsymbol{w} \cdot \boldsymbol{v}_{\mathrm{CM}} \rangle = 0.$$

Pero ¿qué es $\boldsymbol{w} \cdot \boldsymbol{v}_{CM}$? Es

$$\begin{aligned} \boldsymbol{w} \cdot \boldsymbol{v}_{\mathrm{CM}} &= \frac{(\boldsymbol{v}_A - \boldsymbol{v}_B) \cdot (m_A \boldsymbol{v}_A + m_B \boldsymbol{v}_B)}{M} \\ &= \frac{m_A v_A^2 + (m_B - m_A)(\boldsymbol{v}_A \cdot \boldsymbol{v}_B) - m_B v_B^2}{M}. \end{aligned}$$

Por lo tanto, dado que $\langle m_A v_A^2 \rangle = \langle m_B v_B^2 \rangle$, el primero y el último término se anulan en promedio, y nos queda

$$(m_B - m_A) \langle \boldsymbol{v}_A \cdot \boldsymbol{v}_B \rangle = 0.$$

Así, si $m_A \neq m_B$ encontramos que $\langle \boldsymbol{v}_A \cdot \boldsymbol{v}_B \rangle = 0$, y que por lo tanto el movimiento de la molécula como un todo, considerada como una partícula simple de masa M, tiene una energía cinética en promedio igual a $\frac{3}{2}kT$.

Entre paréntesis ¡hemos demostrado también, al mismo tiempo, que la energía cinética media de los movimientos internos de la molécula diatómica, sin considerar el movimiento en conjunto del CM, es $\frac{3}{2}kT$! Porque la energía cinética de las partes de la molécula es $\frac{1}{2} m_A v_A^2 + \frac{1}{2} m_B v_B^2$, cuyo promedio es $\frac{3}{2}kT + \frac{3}{2}kT$, o $3kT$. La energía cinética del movimiento del centro de masas es $\frac{3}{2}kT$; por lo tanto, la energía cinética media de los movimientos de rotación y vibración de los dos átomos en el interior de la molécula es la diferencia, $\frac{3}{2}kT$.

El teorema referente a la energía media del movimiento del CM es general: para cualquier objeto considerado como un todo, con fuerzas presentes o no, para toda dirección independiente de movimiento que haya, la energía cinética media en ese movimiento es $\frac{1}{2}kT$. Estas "direcciones independientes de movimiento" se

llaman a veces *grados de libertad* del sistema. El número de grados de libertad de una molécula compuesta de r átomos es $3r$, ya que cada átomo necesita tres coordenadas para definir su posición. La energía cinética entera de la molécula se puede expresar, ya sea como la suma de las energías cinéticas de los átomos separados, o como la suma de la energía cinética del movimiento del CM más la energía cinética de los movimientos internos. Esta última puede a veces expresarse como una suma de la energía cinética de rotación de la molécula y la energía de vibración, pero ésta es una aproximación. Nuestro teorema, aplicado a la molécula de r átomos, dice que la molécula tendrá en promedio $3rkT/2$ joules de energía cinética, de los cuales $\frac{3}{2}kT$ es energía cinética del movimiento del centro de masas de la molécula entera, y el resto, $\frac{3}{2}(r-1)kT$, es energía cinética interna de vibración y de rotación.

40

Los principios de la mecánica estadística

40-1 La atmósfera exponencial

Hemos discutido algunas de las propiedades de números grandes de átomos que chocan entre sí. El tema se llama teoría cinética, una descripción de la materia desde el punto de vista de choques entre átomos. Fundamentalmente, afirmamos que las propiedades macroscópicas de la materia deberían ser explicables en términos del movimiento de sus partes.

Nos limitaremos por el momento a condiciones de equilibrio térmico, esto es, a una subclase de todos los fenómenos de la naturaleza. Las leyes de la mecánica que se aplican al equilibrio térmico mismo se llaman *mecánica estadística*, y en esta sección queremos familiarizarnos con algunos de los teoremas centrales de este tema.

Ya tenemos uno de los teoremas de la mecánica estadística, es decir, que el valor medio de la energía cinética para cualquier movimiento a la temperatura absoluta T es $\frac{1}{2}kT$, para cada movimiento independiente, es decir, para todos los grados de libertad. Esto nos dice algo acerca de las medias del cuadrado de las velocidades de los átomos. Nuestro objetivo es ahora aprender más acerca de las posiciones de los átomos, descubrir cuántos estarán en lugares diferentes en el equilibrio térmico y también entrar un poco más detalladamente en la distribución de las velocidades. A pesar de que tenemos la media del cuadrado de la velocidad, no sabemos cómo responder a una pregunta como ¿cuántos se mueven tres veces más rápido que la velocidad media cuadrática (raíz de la media del cuadrado de la velocidad)?, o ¿cuántos se mueven con un cuarto de la velocidad media cuadrática?, o ¿es que todos tienen exactamente la misma velocidad?

Así, éstas son las dos preguntas que trataremos de responder: ¿cómo están distribuidas las moléculas en el espacio cuando hay fuerzas que actúan sobre ellas? y ¿cómo están distribuidas en cuanto a la velocidad?

Resulta que las dos preguntas son completamente independientes y que la distribución de velocidades es siempre la misma. Ya hemos recibido una indicación del último hecho cuando encontramos que la energía cinética media es la misma, $\frac{1}{2}kT$ por grado de libertad, actúen las fuerzas que actúen sobre las moléculas. La distribución de velocidades de las moléculas es independiente de las fuerzas, porque las frecuencias de colisión no dependen de las fuerzas.

Comencemos con un ejemplo: la distribución de las moléculas en una atmósfera como la nuestra, pero sin vientos y otros tipos de perturbaciones. Supongan que tenemos una columna de gas que se extiende a una gran altura y en equilibrio térmico —a diferencia de nuestra atmósfera que, como sabemos, se hace más fría a medida que ascendemos—. Podríamos observar que si la temperatura fuera diferente a distintas alturas, demostraríamos la falta de equilibrio, conectando una barra a alguna de las bolas del fondo (figura 40-1), donde éstas escogerían $\frac{1}{2}kT$ de las moléculas y sacudirían, a través de la barra, las bolas en el tope y éstas a su vez sacudirían las moléculas en el tope. Así, por supuesto, la temperatura finalmente llega a ser la misma a todas las alturas en un campo gravitacional.

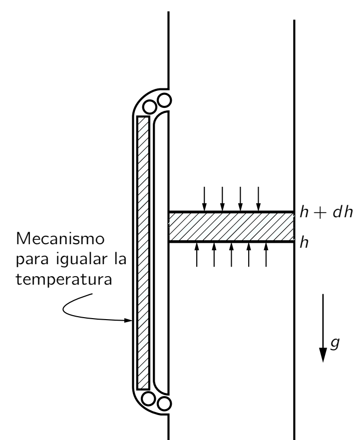


Figura 40-1. La presión a la altura h debe exceder a la presión en $h + dh$ en el peso de gas que interviene.

Si la temperatura es la misma en todas las alturas, el problema es descubrir la ley por la cual la atmósfera se hace más tenue a medida que ascendemos. Si N es el número total de moléculas en un volumen V de un gas a presión P, sabemos que $PV = NkT$ o $P = nkT$, donde $n = N/V$ es el número de moléculas por unidad de volumen. En otras palabras, si conocemos el número de moléculas por unidad de volumen, conoceremos la presión, y viceversa: son proporcionales

entre sí, dado que la temperatura es constante en este problema. Pero la presión no es constante, debe aumentar a medida que la altura se reduce, porque tiene que soportar, por decirlo así, el peso de todo el gas que hay por encima. Ésa es la clave para determinar cómo varía la presión con la altura. Si tomamos un área unitaria a la altura h, la fuerza vertical desde abajo sobre esta área unitaria es la presión P. La fuerza vertical por unidad de área que empuja hacia abajo a una altura $h + dh$ sería la misma, en ausencia de la gravedad, pero aquí no lo es, porque la fuerza desde abajo debe exceder la fuerza desde arriba en el peso del gas en la sección entre h y $h + dh$. Ahora bien, mg es la fuerza de gravedad sobre cada molécula, donde g es la aceleración debida a la gravedad y ndh es el número total de moléculas en la sección unitaria. Así que esto nos da la ecuación diferencial $P_{h+dh} - P_h = dP = -mgn\,dh$. Como $P = nkT$ y T es constante, podemos eliminar ya sea P o n, digamos P, y obtener

$$\frac{dn}{dh} = -\frac{mg}{kT}\,n$$

para la ecuación diferencial, que nos dice cómo disminuye la densidad a medida que aumenta la energía.

Tenemos así una ecuación para la densidad de partículas n, que varía con la altura, pero que tiene una derivada que es proporcional a sí misma. Ahora bien, una función que tiene una derivada proporcional a sí misma es una exponencial y la solución de esta ecuación diferencial es

$$n = n_0 e^{-mgh/kT}. \tag{40.1}$$

Aquí la constante de integración, n_0, es evidentemente la densidad en $h = 0$ (que se puede elegir en cualquier parte), y la densidad decrece exponencialmente con la altura.

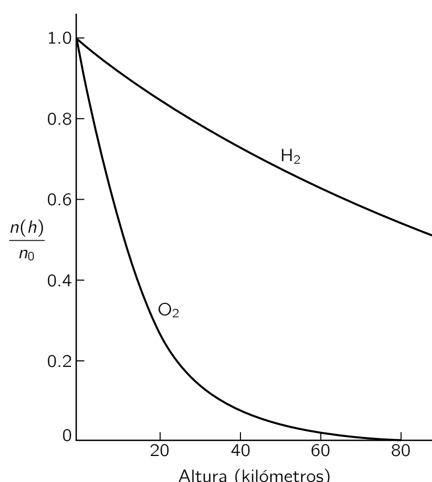


Figura 40-2. La densidad normalizada en función de la altura en el campo gravitacional de la Tierra para oxígeno y para hidrógeno, a temperatura constante.

Nótese que si tenemos diferentes tipos de moléculas con masas distintas, disminuyen con exponenciales diferentes. Las más pesadas disminuirán con la altura más rápidamente que las livianas. Por lo tanto, esperaríamos que, debido a que el oxígeno es más pesado que el nitrógeno, a medida que vamos más y más arriba en una atmósfera, con nitrógeno y oxígeno, la proporción de nitrógeno aumentará. Esto no sucede realmente en nuestra atmósfera, por lo menos a alturas razonables, dado que hay tanta agitación que vuelven a mezclar nuevamente los gases. No es una atmósfera isotérmica. Pero *hay* una tendencia de los materiales más livianos, como el hidrógeno, a predominar a alturas muy grandes en la atmósfera, porque las masas más pequeñas continúan existiendo, mientras que las otras exponenciales se han extinguido todas (figura 40-2).

40-2 La ley de Boltzmann

Notemos aquí el hecho interesante de que el numerador en el exponente de la ecuación (40.1) es la *energía potencial* de un átomo. Por lo tanto, podemos formular también esta ley particular como: la densidad en cualquier punto es proporcional a

$$e^{-\text{la energía potencial de cada átomo}/kT}.$$

Eso puede ser una casualidad, es decir, puede ser válido sólo para este caso particular de un campo gravitacional uniforme. Sin embargo, podemos demostrar que es una proposición más general. Supongamos que existiera algún tipo de fuerza distinta de la gravedad, que actúa sobre las moléculas en un gas. Por ejemplo, las moléculas pueden estar cargadas eléctricamente y pueden ser actuadas por un campo eléctrico u otra carga que las atraiga. O bien, debido a las atracciones mutuas de los átomos entre sí, o por la pared, o por un sólido, o alguna cosa, existe alguna fuerza de atracción que varía con la posición y que actúa sobre todas las moléculas. Supongamos, para simplificar, que todas las moléculas son iguales y que la fuerza actúa sobre cada una individualmente, de manera que la fuerza total sobre una porción de gas sea simplemente el

número de moléculas multiplicado por la fuerza sobre cada una. Para evitar una complicación innecesaria, elijamos un sistema de coordenadas con el eje de las x en la dirección de la fuerza, $\boldsymbol{F}$.

De la misma manera que antes, si tomamos dos planos paralelos en el gas, separados por una distancia dx, la fuerza sobre cada átomo, multiplicada por los n átomos por cm^3 (la generalización del nmg anterior), multiplicado por dx, debe ser equilibrada por la variación de presión: $Fndx = dP = kTdn$. O bien, para poner esta ley en una forma que nos será útil más adelante,

$$F = kT\,\frac{d}{dx}\,(\ln n). \tag{40.2}$$

Observen ahora que $-Fdx$ es el trabajo que realizaríamos al llevar la molécula desde x hasta $x + dx$, y si F proviene de un potencial, es decir, si el trabajo realizado se puede representar por una energía potencial, entonces ésta sería también la diferencia de energía potencial (EP). El diferencial negativo de la energía potencial es el trabajo realizado, $F\,dx$, y encontramos que $d(\ln n) = -d(\text{EP})/kT$, o, después de integrar

$$n = (\text{constante})e^{-\text{EP}/kT}. \tag{40.3}$$

Por lo tanto, lo que habíamos notado para un caso especial resulta ser válido en general. (¿Qué pasa si F no proviene de un potencial? Entonces (40.2) no tiene solución alguna. Puede generarse energía, o perderse por parte de los átomos que se mueven en trayectorias cíclicas para las que el trabajo realizado no es cero, y no se puede mantener ningún equilibrio. El equilibrio térmico no puede existir cuando las fuerzas externas sobre los átomos no son conservativas.) La ecuación (40.3), conocida como *ley de Boltzmann*, es otro de los principios de la mecánica estadística: que la probabilidad de encontrar moléculas en un arreglo espacial dado varía exponencialmente con menos la energía potencial de ese arreglo, dividida entre kT.

Esto podría indicarnos entonces la distribución de las moléculas: supongan que tuviéramos un ion positivo en un líquido, que atrae los iones negativos a su alrededor, ¿cuántos estarían a diferentes distancias? Si se conoce la energía potencial en función de la distancia, la proporción de ellos a diferentes distancias está dada por esta ley, y así sucesivamente, para muchas aplicaciones.

40-3 Evaporación de un líquido

En mecánica estadística más avanzada, uno trata de resolver el siguiente problema importante. Consideremos un conjunto de moléculas que se atraen entre sí y supongamos que la fuerza entre dos cualesquiera, digamos i y j, depende sólo de su separación r_{ij} y es posible representarla como la derivada de una función potencial $V(r_{ij})$. En la figura 40-3 se muestra una forma que podría tener tal función. Para $r > r_0$ la energía decrece a medida que las moléculas se acercan, porque se atraen, y luego la energía crece muy abruptamente a medida que se acercan aún más, porque se repelen fuertemente, lo que es característico de la manera como se comportan las moléculas, en términos generales.

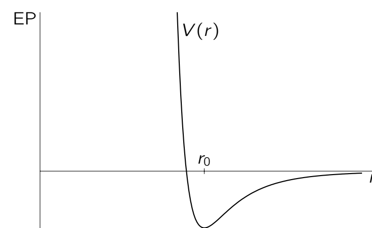


Figura 40-3. Una función de energía potencial para dos moléculas, que depende sólo de su separación.

Supongamos ahora que tenemos una caja llena de estas moléculas, y nos gustaría saber cómo se disponen entre sí, en promedio. La respuesta es $e^{-\text{EP}/kT}$. La energía potencial total, en este caso, sería la suma sobre todos los pares, suponiendo que las fuerzas están todas en pares (podría haber fuerzas de tres cuerpos en cosas más complicadas, pero en la electricidad, por ejemplo, la energía potencial es enteramente en pares). Entonces la probabilidad de encontrar moléculas en cualquier combinación particular de los r_{ij} será proporcional a

$$\exp\Big[-\sum_{i,j} V(r_{ij})/kT\Big].$$

Ahora bien, si la temperatura es muy alta, de modo que $kT \gg |V(r_0)|$, el exponente será relativamente pequeño en casi todas partes, y la probabilidad de

encontrar una molécula es casi independiente de la posición. Tomemos el caso de sólo dos moléculas; entonces $e^{-\text{EP}/kT}$ sería la probabilidad de encontrarlas a varias distancias mutuas r. Claramente, donde el potencial se hace más negativo, la probabilidad es mayor y donde el potencial va a infinito la probabilidad es cero, lo que ocurre para distancias muy pequeñas. Eso significa que para tales átomos en un gas no hay ninguna posibilidad de que estén uno encima del otro, ya que se repelen muy fuertemente. Pero hay una probabilidad mayor de encontrarlos *por unidad de volumen* en el punto r_0 que en cualquier otro punto. En cuánto mayor depende de la temperatura. Si la temperatura es muy alta comparada con la diferencia en energía entre $r = r_0$ y $r = \infty$, la exponencial es siempre casi unitaria. En este caso, donde la energía cinética media (alrededor de kT) excede grandemente a la energía potencial, las fuerzas no importan mucho. Pero a medida que la temperatura desciende, la probabilidad de encontrar las moléculas a la distancia preferida r_0 aumenta gradualmente respecto a la probabilidad de encontrarlas en cualquier otra parte y, de hecho, si kT es mucho menor que $|V(r_0)|$ tenemos un exponente positivo relativamente grande en esa vecindad. En otras palabras, en un volumen dado tienen *mucha más* probabilidad de estar a la distancia de energía mínima que muy separadas. A medida que la temperatura disminuye, los átomos se juntan, se amontonan, y se reducen a líquidos, sólidos, y moléculas, y cuando se calientan se evaporan.

Los requisitos para determinar exactamente cómo se evaporan las cosas, cómo suceden exactamente las cosas en una circunstancia dada, comprenden lo siguiente. Primero descubrir la ley correcta de la fuerza molecular $V(r)$, que debe provenir de alguna otra cosa, digamos la mecánica cuántica o el experimento. Pero, dada la ley de fuerza entre las moléculas, descubrir qué van a hacer un billón de moléculas consiste meramente en estudiar la función $e^{-\sum V_{ij}/kT}$. Es bastante sorprendente ya que se trata de una función tan simple y de una idea tan sencilla, dado el potencial, que la labor sea tan *enormemente complicada*, la dificultad es el tremendo número de variables.

A pesar de tales dificultades, el tema es emocionante e interesante. A menudo se le denomina un ejemplo del "problema de muchos cuerpos" y ha sido realmente una cosa muy interesante. En esta simple fórmula deben estar contenidos todos los detalles, por ejemplo, sobre la solidificación de un gas, o las formas de los cristales que un sólido puede adoptar, y la gente ha tratado de exprimirla, pero las dificultades matemáticas son muy grandes, no en escribir la ley, sino en tratar con un número tan grande de variables.

Ésa es entonces la distribución de partículas en el espacio. Ése es el final de la mecánica estadística clásica, hablando en forma práctica, porque si conocemos las fuerzas podemos, en principio, encontrar la distribución en el espacio, y la distribución de las velocidades es algo que podemos determinar de una vez por todas y no algo que sea diferente para los distintos casos. Los grandes problemas están en obtener información particular a partir de nuestra solución formal y éste es el objeto principal de la mecánica estadística clásica.

40-4 La distribución de las velocidades moleculares

Ahora pasamos a discutir la distribución de velocidades, porque a veces es interesante o útil saber cuántas se mueven a velocidades diferentes. Para realizar eso, podemos hacer uso de lo que hemos descubierto en relación con el gas en la atmósfera. Lo consideramos un gas perfecto, como ya lo hemos supuesto al escribir la energía potencial, despreciando la energía de atracción mutua de los átomos. La única energía potencial que incluimos en nuestro primer ejemplo fue la gravedad. Tendríamos, por supuesto, algo más complicado si existieran fuerzas entre los átomos. Así, suponemos que no hay fuerzas entre los átomos y por un momento despreciamos también las colisiones, volviendo más tarde a la justificación de esto. Ahora bien, vimos que hay menos moléculas a la altura h que a la altura 0; de acuerdo con la fórmula (40.1) decrecen exponencialmente con la altura. ¿Cómo puede haber menos a alturas mayores? ¿Después de todo, no llegarán a h todas las moléculas que se mueven hacia arriba desde la altura 0?

¡No!, porque algunas de las que se mueven hacia arriba en 0 van demasiado lento y no pueden escalar la montaña de potencial hasta h. Con esta clave podemos calcular cuántas deben estar moviéndose a diversas velocidades, porque según (40.1) sabemos *cuántas* se mueven con una velocidad menor que la suficiente para escalar una distancia h dada. Ésas son precisamente las que explican el hecho de que la densidad en h es menor que en 0.

Pongamos ahora esta idea en una forma un poco más precisa: contemos cuántas moléculas pasan desde abajo hacia encima del plano $h = 0$ (al llamarlo altura $= 0$, no queremos decir que el piso esté ahí; es sólo una referencia conveniente y existe gas a h negativo). Estas moléculas de gas se mueven en todas direcciones, pero algunas se mueven a través del plano, y en cada instante un cierto número por segundo pasa a través del plano desde abajo hacia arriba, con diferentes velocidades. Notamos ahora lo siguiente: si llamamos u la velocidad que se necesita precisamente para alcanzar la altura h (energía cinética $mu^2/2 = mgh$), entonces el número de moléculas por segundo que pase hacia arriba a través del plano inferior en una dirección vertical con una componente de la velocidad mayor que u es exactamente el mismo que el número que pasa a través del plano superior con *cualquier* velocidad hacia arriba. Las moléculas cuya velocidad vertical no excede u no pueden atravesar el plano superior. Por lo tanto, vemos

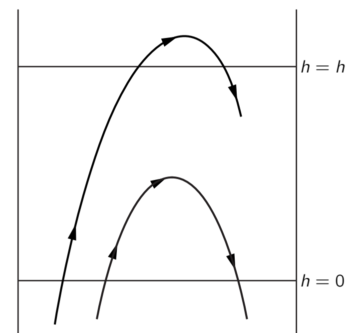


Figura 40-4. Sólo las moléculas que se mueven hacia arriba en $h = 0$ con velocidad suficiente pueden llegar a la altura h.

$$\text{Núm. que pasa por } h = 0 \text{ con } v_z > u = \text{núm. que pasa por } h = h \text{ con } v_z > 0.$$

Pero el número que pasa por h con cualquier velocidad mayor que 0 es menor que el número que pasa por la altura inferior con cualquier velocidad mayor que 0, porque el número de átomos es mayor; eso es todo lo que necesitamos. Ya sabemos que la distribución de velocidades es la misma, según el razonamiento que hicimos antes respecto a la constancia de la temperatura a través de toda la atmósfera. Así, como las distribuciones de velocidades son las mismas, y que además existen *más átomos* más abajo, claramente el número $n_{>0}(h)$, que pasa con velocidad positiva a la altura h y el número $n_{>0}(0)$ que pasa con velocidad positiva a la altura 0, están a la misma razón que las densidades a las dos alturas, que es $e^{-mgh/kT}$. Pero $n_{>0}(h) = n_{>u}(0)$ y, por lo tanto, encontramos que

$$\frac{n_{>u}(0)}{n_{>0}(0)} = e^{-mgh/kT} = e^{-mu^2/2kT},$$

ya que $\frac{1}{2}mu^2 = mgh$. Así, en palabras, el número de moléculas por unidad de área por segundo que pasan la altura 0 con un componente z de la velocidad mayor que u es $e^{-mu^2/2kT}$ multiplicado por el número total que pasa por el plano con velocidad mayor que cero.

Ahora bien, esto no es sólo válido a la altura 0 elegida arbitrariamente, pero, por supuesto, es válido para cualquier otra altura y, por lo tanto, ¡las distribuciones de velocidades *son* todas iguales! (En el enunciado final no interviene la altura h que aparece solamente en el razonamiento intermedio.) El resultado es una proposición general que nos da la distribución de velocidades. Nos dice que si perforamos un pequeño agujero en el costado de una cañería de gas, un agujero muy chiquito, tal que las colisiones sean pocas y distanciadas, es decir, estén más distanciadas que el diámetro del agujero, entonces las partículas que salen tendrán velocidades diferentes, pero la fracción de partículas que sale a una velocidad mayor que u es $e^{-mu^2/2kT}$.

Volvamos ahora al hecho de haber despreciado los choques: ¿por qué esto no tiene importancia? Podríamos haber seguido el mismo razonamiento, no con una altura finita h, sino que con una altura infinitesimal h, que es tan pequeña que no hubiera lugar para colisiones entre 0 y h. Pero eso no fue necesario, el razonamiento está evidentemente basado sobre un análisis de las energías comprendidas, la conservación de la energía, y en las colisiones que ocurren hay un intercambio de energías entre las moléculas. Sin embargo, no nos preocupamos realmente, si seguimos la misma molécula, si la energía se intercambia meramente con otra molécula. Resulta así que, aun si el problema

se analiza más cuidadosamente (y es más difícil, naturalmente, hacer un trabajo más riguroso), todavía no introduce diferencia en el resultado.

Es interesante que la distribución de velocidades que hemos encontrado sea precisamente

$$n_{>u} \propto e^{-\text{energía cinética}/kT}. \tag{40.4}$$

Esta manera de describir la distribución de velocidades, dando el número de moléculas que pasa por un área dada, con una cierta componente z mínima, no es la manera más conveniente para dar la distribución de velocidad. Por ejemplo, dentro del gas, uno desea más a menudo saber cuántas moléculas se mueven con una componente z de la velocidad entre dos valores dados, y esto por supuesto no está dado directamente por la ecuación (40.4). Nos gustaría establecer nuestro resultado en la forma más convencional, a pesar de que lo que ya hemos escrito es bastante general. *Noten que no es posible decir que cualquier molécula tenga exactamente cierta velocidad establecida*; ninguna de ellas tiene una velocidad *exactamente igual* a 1.7962899173 metros por segundo. Así, para hacer una afirmación que tenga sentido, tenemos que preguntar cuántas se encuentran en algún *intervalo* de velocidades. Debemos decir cuántas tienen velocidades entre 1.796 y 1.797, y así sucesivamente. En términos matemáticos, sea $f(u)\,du$ la fracción de todas las moléculas que tienen velocidades entre u y $u + du$, o lo que es lo mismo (si du es infinitesimal), todas las que tienen una velocidad u en intervalo du. La figura 40-5 muestra una forma posible para la función $f(u)$, y la parte sombreada de ancho du y altura media $f(u)$ representa esta fracción $f(u)\,du$. Esto es, el cociente entre el área sombreada y el área total de la curva es la proporción relativa de moléculas con velocidad u dentro de du. Si definimos $f(u)$ de modo que la fracción que tiene una velocidad en ese intervalo esté dada directamente por el área sombreada, entonces el área total debe ser 100 % de ellas, eso es

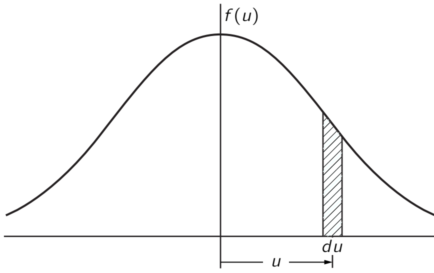


Figura 40-5. Una función de distribución de velocidades. El área sombreada es $f(u)\,du$, representa la fracción de las partículas que tienen velocidades dentro del intervalo du en torno a u.

$$\int_{-\infty}^{\infty} f(u)\,du = 1. \tag{40.5}$$

Ahora bien, tenemos que obtener solamente esta distribución, comparándola con el teorema que hemos deducido anteriormente. Preguntamos primero: ¿cuál es el número de moléculas que atraviesa un área por segundo con una velocidad mayor que u, expresado en términos de $f(u)$? Al comienzo podríamos pensar que sea simplemente la integral de $\int_u^\infty f(u)du$, pero no lo es, porque queremos el número que atraviesa el área *por segundo*. Las más rápidas pasan más a menudo, por decir así, que las más lentas, y para expresar cuántas pasan, deben multiplicar por la velocidad; discutimos esto en el capítulo anterior al hablar del número de colisiones. En un tiempo dado t, el número total que atraviesa la superficie es el de todas aquellas que han podido llegar a la superficie y el número que llega viene desde una distancia ut. Así, pues, el número de moléculas que llega no es simplemente el número que está allí, sino el número que está allí por unidad de volumen, multiplicado por la distancia que barren al moverse hacia el área a través de la cual se supone que van, y esa distancia es proporcional a u. Por lo tanto, necesitamos la integral de u *multiplicado* por $f(u)du$, una integral infinita, con un límite inferior u, y esto debe ser lo mismo que hemos encontrado antes, es decir, $e^{-mu^2/2kT}$; con una constante de proporcionalidad que obtendremos más adelante:

$$\int_u^\infty uf(u)\,du = \text{constante} \cdot e^{-mu^2/2kT}. \tag{40.6}$$

Ahora bien, si derivamos la integral respecto a u, obtenemos lo que está dentro de la integral, es decir, el integrando (con un signo menos, ya que u es el límite inferior), y si derivamos el otro miembro, obtenemos u veces la misma exponencial (y algunas constantes). Las u se simplifican y encontramos

$$f(u)\,du = Ce^{-mu^2/2kT}\,du. \tag{40.7}$$

Retenemos los du en ambos miembros para recordar que se trata de una *distribución*, que dice cuál es la proporción para una velocidad entre u y $u + du$.

La constante C debe determinarse de manera que la integral sea la unidad, de acuerdo a la ecuación (40.5). Ahora podemos demostrar* que

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

Usando este hecho, es fácil encontrar que $C = \sqrt{m/2\pi kT}$.

Como la velocidad y el momentum son proporcionales, podemos decir que la distribución de momenta es también proporcional a $e^{-\mathrm{EC}/kT}$ por unidad de intervalo de momentum. Este teorema es válido también en relatividad, si está en términos del momentum, mientras que no lo es si está en términos de la velocidad, así que es mejor aprenderlo en términos de momentum en vez de velocidad:

$$f(p)\,dp = Ce^{-\mathrm{EC}/kT}\,dp. \tag{40.8}$$

Encontramos así que las probabilidades de diferentes condiciones de energía, cinética y potencial, están ambas dadas por $e^{-\text{energía}/kT}$, algo muy fácil de recordar y una proposición bastante bella.

Hasta ahora tenemos, por supuesto, sólo la distribución de las velocidades "verticalmente". Podríamos preguntar: ¿cuál es la probabilidad de que una molécula se mueva en otra dirección? Por supuesto, estas distribuciones están conectadas y se puede obtener la distribución completa a partir de la que tenemos, porque la distribución completa depende solamente del cuadrado del módulo de la velocidad, no de la componente z. Debe ser algo que sea independiente de la dirección y hay sólo una función que interviene, la probabilidad de diferentes módulos. Tenemos la distribución de la componente z y por lo tanto podemos obtener la distribución de otras componentes a partir de ella. El resultado es que la probabilidad es todavía proporcional a $e^{-\mathrm{EC}/kT}$, pero ahora la energía cinética comprende tres partes, $mv_x^2/2$, $mv_y^2/2$ y $mv_z^2/2$, sumadas en el exponente. O podemos escribirlo como un producto:

$$f(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z \propto e^{-mv_x^2/2kT} \cdot e^{-mv_y^2/2kT} \cdot e^{-mv_z^2/2kT}\,dv_x\,dv_y\,dv_z. \tag{40.9}$$

Pueden ver que esta fórmula debe ser correcta, porque, primero, es una función de v^2 solamente, como se requiere, y segundo, las probabilidades de los diversos valores de v_z obtenidas por integración sobre todos los v_x y v_y es precisamente (40.7). ¡Así esta sola función (40.9) puede hacer ambas cosas!

40-5 Calores específicos de gases

Examinemos ahora algunas maneras de comprobar la teoría, y ver hasta qué punto ha tenido éxito la teoría clásica de gases. Vimos anteriormente que si U es la energía interna de N moléculas, entonces $PV = NkT = (\gamma - 1)U$ es cierto, algunas veces, para algunos gases, quizás. Si es un gas monoatómico, sabemos que esto es igual a $\frac{2}{3}$ de la energía cinética del movimiento del centro de masa de los átomos. Si es un gas monoatómico, la energía cinética es igual a la

---

* Para obtener el valor de la integral, hagamos

$$I = \int_{-\infty}^{\infty} e^{-x^2}\,dx.$$

Entonces,

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2}\,dx \cdot \int_{-\infty}^{\infty} e^{-y^2}\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dy\,dx,$$

que es una integral doble sobre todo el plano xy. Pero ésta puede escribirse también en coordenadas polares como

$$I^2 = \int_0^{\infty} e^{-r^2} \cdot 2\pi r\,dr = \pi \int_0^{\infty} e^{-t}\,dt = \pi.$$

energía interna y, por lo tanto, $\gamma - 1 = \frac{2}{3}$. Pero supongamos que es, digamos, una molécula más complicada, que puede rotar en torno a sí misma y vibrar, y supongamos, resulta ser cierto de acuerdo a la mecánica clásica, que las energías de los movimientos internos son también proporcionales a kT. Entonces, a una temperatura dada, además de la energía cinética $\frac{3}{2}kT$, tiene energía interna de vibración o de rotación. Así pues, el total U no sólo incluye la energía cinética interna, sino también la energía de rotación, y obtenemos un valor diferente de γ. Técnicamente, la mejor manera de medir γ es midiendo el calor específico, que es la variación de energía con la temperatura. Volveremos a este método más adelante. Para nuestros fines presentes, podemos suponer que γ se encuentra experimentalmente a partir de la curva PV^γ para la compresión adiabática.

Hagamos un cálculo de γ para algunos casos. Primero, para un gas monoatómico U es la energía total, igual a la energía cinética, y ya sabemos que γ debería ser $\frac{5}{3}$. Como un gas diatómico, podemos tomar, como ejemplo, oxígeno, yoduro de hidrógeno, hidrógeno, etc., y suponer que el gas biatómico puede representarse como dos átomos unidos por algún tipo de fuerza como la de la figura 40-3. Podemos suponer también, y resulta ser cierto, que a las temperaturas que son de interés para el gas diatómico, los pares de átomos tienden fuertemente a estar separados por r_0, la distancia de potencial mínimo. Si esto no fuera verdad, si la probabilidad no variara lo suficientemente fuerte como para lograr que la gran mayoría esté situada cerca del fondo, tendríamos que recordar que el gas oxígeno sería una mezcla de O_2 y átomos simples de oxígeno en una proporción significativa. Sabemos, en efecto, que hay muy pocos átomos simples de oxígeno, lo que significa que el mínimo de la energía potencial es mucho mayor en valor absoluto que kT, como hemos visto. Como están fuertemente acumulados alrededor de r_0, la única parte de la curva que se necesita es la parte cerca del mínimo, que puede aproximarse por una parábola. Un potencial parabólico implica un oscilador armónico y, de hecho, con una aproximación excelente, la molécula de oxígeno se puede representar como dos átomos conectados por un resorte.

¿Cuál es ahora la energía total de esta molécula a la temperatura T? Sabemos que para cada uno de los átomos, cada una de las energías cinéticas debería ser $\frac{3}{2}kT$; entonces, la energía cinética de ambos es $\frac{3}{2}kT + \frac{3}{2}kT$. Podemos poner esto también de una manera diferente: el mismo $\frac{3}{2}$ más $\frac{3}{2}$ puede considerarse como la energía cinética del centro de masa $(\frac{3}{2})$, energía cinética de rotación $(\frac{2}{2})$ y energía cinética de vibración $(\frac{1}{2})$. Sabemos que la energía cinética de vibración es $\frac{1}{2}$, ya que hay sólo una dimensión en juego, de cada grado de libertad tiene $\frac{1}{2}kT$. Respecto a la rotación, ella puede rotar alrededor de dos ejes cualesquiera, por lo que hay dos movimientos independientes. Suponemos que los átomos son una especie de puntos y no pueden girar alrededor de la línea que los une; esto es algo que hay que tener en mente, porque si obtenemos algún desacuerdo, quizás esté ahí el problema. Pero tenemos una cosa más, la energía *potencial* de vibración, ¿cuánto es? En un oscilador armónico, la energía cinética media y la energía potencial media son iguales y, por lo tanto, la energía potencial de vibración es también $\frac{1}{2}kT$. El gran total de la energía es $U = \frac{7}{2}kT$, o kT es $\frac{2}{7}U$ por átomo. Eso significa, entonces, que γ es $\frac{9}{7}$ en lugar de $\frac{5}{3}$, es decir, $\gamma = 1.286$.

Tabla 40-1

Valores del cociente de los calores específicos, γ, para diversos gases

Gas	T (°C)	γ
He	−180	1.660
Kr	19	1.68
Ar	15	1.668
H_2	100	1.404
O_2	100	1.399
HI	100	1.40
Br_2	300	1.32
I_2	185	1.30
NH_3	15	1.310
C_2H_6	15	1.22

Podemos comparar estos números con los valores pertinentes medidos mostrados en la tabla 40-1. Mirando primero el helio, que es un gas monoatómico, encontramos muy cercanamente $\frac{5}{3}$, y el error es probablemente experimental, a pesar de que a temperatura tan baja puede haber alguna fuerza entre los átomos. Criptón y argón, ambos monoatómicos, concuerdan también, dentro de la precisión del experimento.

Pasamos a los gases biatómicos y encontramos hidrógeno con 1.404, lo que no está de acuerdo con la teoría, 1.286. Oxígeno, 1.399, es muy similar, pero de nuevo en desacuerdo. Yoduro de hidrógeno es nuevamente similar a 1.40. Empieza a parecer como si la respuesta correcta fuera 1.40, pero no lo es, porque si seguimos mirando en el bromo vemos 1.32, y en el yodo vemos 1.30. Como 1.30 está razonablemente cerca de 1.286, se puede decir que el yodo concuerda bastante bien, pero el oxígeno está lejos. Tenemos aquí, por lo tanto, un dilema.

Lo tenemos correcto para una molécula, no lo tenemos correcto para otra molécula y necesitaríamos ser bastante ingeniosos para explicar ambas circunstancias.

Sigamos mirando una molécula todavía más complicada, con gran número de partes, por ejemplo, C_2H_6, que es etano. Tiene ocho átomos diferentes, y todos están vibrando y rotando en diversas combinaciones, de modo que el monto total de la energía interna debe ser un número enorme de kT, por lo menos $12kT$ sólo para la energía cinética, y $\gamma - 1$ debe estar muy cerca de cero, o γ es casi exactamente 1. En efecto, *es* menor, pero 1.22 no es tanto menor, y es superior a $1\frac{1}{12}$ calculado a partir de la energía cinética solamente, ¡y esto es simplemente incomprensible!

Más aún, el misterio completo es profundo, porque la molécula diatómica no puede hacerse rígida hasta el límite. Aun si hiciéramos los acoplamientos indefinidamente rígidos, a pesar de que no podría vibrar mucho, se mantendría de todos modos vibrando. La energía de vibración interna es todavía kT, dado que no depende de la intensidad de acoplamiento. Pero si pudiéramos imaginar rigidez *absoluta*, deteniendo toda vibración para eliminar una variable, entonces obtendríamos $U = \frac{5}{2}kT$ y $\gamma = 1.40$ para el caso diatómico. Esto parece bien para H_2 u O_2. ¡Por otro lado, todavía tendríamos problemas, porque γ varía, tanto para hidrógeno como oxígeno, con la temperatura! De los valores medidos mostrados en la figura 40-6 vemos que para H_2, γ varía desde alrededor desde 1.6 a $-185°$C hasta 1.3 a 2 000°C. La variación es más sustancial en el caso del hidrógeno que para el oxígeno: sin embargo, aun para el oxígeno, γ tiende definidamente a crecer, a medida que descendemos en temperatura.

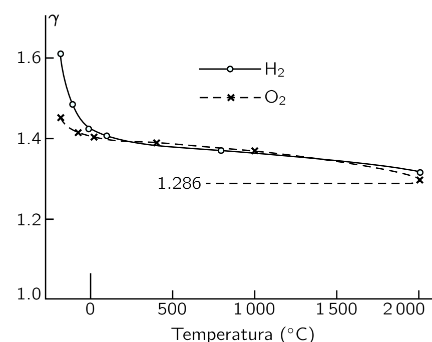


Figura 40-6. Valores experimentales de γ en función de la temperatura para hidrógeno y oxígeno. La teoría clásica predice $\gamma = 1.286$, independientemente de la temperatura.

40-6 El fracaso de la física clásica

Así, en suma, podemos decir que tenemos algunas dificultades. Podríamos probar alguna ley de fuerza diferente a la de un resorte, pero resulta que cualquier otra cosa hará solamente γ mayor. Si incluimos más formas de energía, γ se acerca más a la unidad contradiciendo los hechos. Todas las cosas de la teoría clásica en que uno pudiera pensar, sólo empeorarán la situación. El hecho es que hay electrones en cada átomo, y sabemos, a partir de sus espectros que hay movimientos internos: cada uno de los electrones debería tener por los menos $\frac{1}{2}kT$ de energía cinética y algo para la energía potencial, de modo que cuando éstos se agregan, γ se hace aún menor. Esto es ridículo. Está mal.

La primera gran publicación sobre la teoría dinámica de los gases fue de Maxwell en 1859. Sobre la base de las ideas que hemos estado discutiendo pudo explicar en forma precisa muchísimas relaciones conocidas, como la ley de Boyle, la teoría de difusión, la viscosidad de los gases, y cosas sobre las que hablaremos en el próximo capítulo. Él enumeró todos estos grandes éxitos en un resumen final, y al concluir dijo: "Finalmente, estableciendo una relación necesaria entre los movimientos de traslación y rotación (está hablando del teorema $\frac{1}{2}kT$) de todas las partículas no esféricas, probamos que un sistema de tales partículas no podría posiblemente satisfacer las relaciones conocidas entre los dos calores específicos". Se está refiriendo a γ (que, como veremos más adelante, está relacionado a dos formas de medir calor específico), y dice que sabemos que no podemos obtener la respuesta correcta.

Diez años más tarde, en una clase, dijo: "He expuesto ahora ante ustedes lo que considero la dificultad más grande encontrada hasta ahora por la teoría molecular". Estas palabras representan el primer descubrimiento de que las leyes de la física clásica estaban erradas. Ésta fue la primera indicación de que había algo fundamentalmente imposible, porque un teorema demostrado rigurosamente no concordaba con el experimento. Alrededor de 1905, Sir James Hopwood Jeans y Lord Rayleigh (John William Strutt) hablaron de nuevo sobre este enigma. A menudo uno escucha decir que los físicos en la última parte del siglo XIX pensaban que conocían todas las leyes físicas significativas y que todo lo que tenían que hacer era calcular más cifras decimales. Alguien puede haber dicho esto alguna vez, y otros lo copiaron. Pero una lectura minuciosa de la literatura de ese tiempo muestra que todos estaban preocupados de algo. Jeans decía respecto a

ese enigma que es un fenómeno muy misterioso y parece como que si a medida que la temperatura baja, ciertos tipos de movimiento "se congelan".

Si pudiéramos suponer que el movimiento vibratorio, por decir, no existe a temperatura baja y existe a temperatura alta, entonces podríamos imaginar que un gas pueda existir a una temperatura suficientemente baja como para que el movimiento de vibración no ocurra, de modo que $\gamma = 1.40$, o a una temperatura mayor, a la cual comienza a aparecer de modo que γ disminuye. Lo mismo se podría argumentar para la rotación. Si podemos eliminar la rotación, digamos que "se congela" a temperatura suficientemente baja, entonces podemos entender el hecho que el γ del hidrógeno se aproxima a 1.66 a medida que bajamos en temperatura. ¿Cómo podemos entender tal fenómeno? Por supuesto, el que estos movimientos "se congelen" no se puede entender con la mecánica clásica. Fue comprendido solamente cuando se descubrió la mecánica cuántica.

Sin demostración podemos enunciar los resultados de la teoría cuántica para la mecánica estadística. Recordemos que, de acuerdo con la mecánica cuántica, un sistema que está ligado por un potencial, para las vibraciones, por ejemplo, tendrá un conjunto discreto de niveles energéticos, es decir, estados de diferente energía. El problema es ahora: ¿cómo se modifica la mecánica estadística de acuerdo con la teoría cuántica? ¡Resulta, y es muy interesante, que a pesar de que la mayoría de los problemas son más difíciles en la mecánica cuántica que en la mecánica clásica, los problemas en la mecánica estadística son mucho más fáciles en la teoría cuántica! El resultado sencillo que tenemos en la mecánica clásica, que $n = n_0 e^{-\text{energía}/kT}$, pasa a ser el siguiente teorema muy importante: si las energías del conjunto de estados moleculares se llaman, digamos, $E_0, E_1, E_2, \ldots, E_i, \ldots$, entonces en el equilibrio térmico la probabilidad de encontrar una molécula en un estado particular que tiene energía E_i es proporcional a $e^{-E_i/kT}$. Eso da la probabilidad de estar en diversos estados. En otras palabras, la posibilidad relativa, la probabilidad de estar en el estado E_1 relativa a la posibilidad de estar en el estado E_0 es:

$$\frac{P_1}{P_0} = \frac{e^{-E_1/kT}}{e^{-E_0/kT}}, \tag{40.10}$$

que es, por supuesto, lo mismo que

$$n_1 = n_0 e^{-(E_1 - E_0)/kT}, \tag{40.11}$$

ya que $P_1 = n_1/N$ y $P_0 = n_0/N$. Así es menos probable estar en un estado energético más alto que en uno más bajo. La razón del número de átomos en el estado superior al número en el estado más bajo es e elevado a la potencia (menos la diferencia de energía, dividido entre kT) —una proposición muy simple—.

Resulta ahora que para un oscilador armónico, los niveles energéticoss están igualmente espaciados. Llamando $E_0 = 0$ la energía más baja (en realidad no es cero, es un poco diferente, pero no tiene importancia si ajustamos todas las energías con una constante), la primera es, entonces, $E_1 = \hbar\omega$, y la segunda es $2\hbar\omega$, y la tercera es $3\hbar\omega$, y así sucesivamente.

Veamos ahora qué sucede. Supongan que estudiamos las vibraciones de una molécula diatómica, que aproximamos con un oscilador armónico. Preguntemos cuál es la probabilidad relativa de encontrar una molécula en el estado E_1 en vez de en el estado E_0. La respuesta es que la probabilidad de encontrarla en el estado E_1, relativa a la de encontrarla en el estado E_0, disminuye como $e^{-\hbar\omega/kT}$. Supongamos ahora que kT es mucho menor que $\hbar\omega$ y tendremos una situación a baja temperatura. Entonces la probabilidad de que esté en el estado E_1 es extremadamente pequeña. Prácticamente todos los átomos están en el estado E_0. Si cambiamos la temperatura, pero todavía la mantenemos muy baja, entonces la probabilidad de que esté en el estado $E_1 = \hbar\omega$ permanece infinitesimal: la energía del oscilador permanece cerca de cero; no cambia con la temperatura mientras la temperatura sea mucho menor que $\hbar\omega$. Todos los osciladores están en el estado más bajo, y sus movimientos están efectivamente "congelados"; *no hay ninguna contribución suya al calor específico.* Podemos juzgar, entonces, a partir

de la tabla 40-1 que a 100°C, que es 373 grados absolutos, kT es mucho menor que la energía vibracional en las moléculas de oxígeno o hidrógeno, pero no así en la molécula de yodo. La razón de la diferencia es que el átomo de yodo es muy pesado, comparado con el hidrógeno, y a pesar de que las fuerzas en el yodo y en el hidrógeno pueden ser comparables, la molécula de yodo es tan pesada que la frecuencia natural de vibración es muy baja comparada con la frecuencia natural del hidrógeno. Con $\hbar\omega$ mayor que kT a la temperatura ambiente para el hidrógeno, pero menor para el yodo, solamente el último, el yodo, exhibe la energía clásica de vibración. A medida que aumentamos la temperatura de un gas, partiendo de un valor muy bajo de T, con casi todas las moléculas en su estado más bajo, ellas comienzan gradualmente a tener una probabilidad apreciable de estar en el segundo estado, y luego en el estado siguiente, y así sucesivamente. Cuando la probabilidad es apreciable para muchos estados, el comportamiento del gas se aproxima al dado por la física clásica, porque los estados cuantizados se hacen casi indistinguibles de un continuo de energías, y el sistema puede tener casi cualquier energía. Por lo tanto, a medida que aumenta la temperatura, deberíamos obtener de nuevo los resultados de la física clásica, como realmente parece ser el caso en la figura 40-6. Es posible demostrar, en la misma forma que los estados rotacionales de los átomos están también cuantizados, pero los estados están tanto más próximos, que en situaciones ordinarias kT es mayor que el espaciamiento. Entonces se excitan muchos niveles y la energía cinética de rotación en el sistema participa en la forma clásica. El único ejemplo, para el cual esto no es tan cierto a la temperatura ambiente, es el hidrógeno.

Ésta es la primera vez que realmente hemos deducido, por comparación con el experimento, que había algo errado en la física clásica y hemos buscado una solución de la dificultad en la mecánica cuántica de manera muy similar a como fue hecho originalmente. Pasaron 30o 40 años antes de que se descubriera la dificultad siguiente y esto nuevamente tuvo que ver con la mecánica estadística, pero esta vez la mecánica de un gas de fotones. Este problema fue resuelto por Planck en los primeros años del siglo XX.

41

El movimiento browniano

41-1 Equipartición de la energía

El movimiento browniano fue descubierto en 1827 por el botánico Robert Brown. Mientras estudiaba la vida microscópica, notó pequeñas partículas de polen de plantas que zigzagueaban por el líquido que estaba examinando al microscopio, y fue lo suficientemente inteligente como para darse cuenta de que las mismas no eran vivientes, sino trocitos de suciedad moviéndose por el agua. De hecho, ayudó a demostrar que esto no tenía nada que ver con la vida, tomando de la tierra un viejo pedazo de cuarzo en el que había un poco de agua atrapada. Debía haber estado atrapada por millones y millones de años y, sin embargo, vio en ella el mismo movimiento. Lo que se ve es partículas muy pequeñitas zigzagueando todo el tiempo.

Más tarde se probó que éste era uno de los efectos del *movimiento moleculare*; podemos entenderlo cualitativamente imaginando una gran pelota en una cancha vista desde gran distancia, con mucha gente abajo tirando la pelota en diversas direcciones. No podemos ver la gente porque imaginamos estar demasiado lejos, pero podemos ver la pelota y notamos que se mueve bastante irregularmente de un lado a otro. También sabemos, por los teoremas discutidos en capítulos precedentes, que la energía cinética media de una pequeña partícula en suspensión en un líquido o un gas será $\frac{3}{2}kT$ aunque sea muy pesada en comparación con una molécula. Si es muy pesada, las velocidades son relativamente bajas, pero en la realidad resulta que la velocidad no es tan baja. De hecho, no podemos ver muy fácilmente la velocidad de esas partículas porque, aunque la energía cinética media es $\frac{3}{2}kT$, que representa una velocidad de un milímetro por segundo más o menos para el caso de un objeto de uno o dos micrones de diámetro, esto es muy difícil de ver aun con un microscopio, porque la partícula está continuamente cambiando de dirección y nunca llega a ninguna parte. Al final de este capítulo discutiremos hasta dónde llega. El primero en resolver este problema fue Einstein a comienzos del siglo XX.

Entre paréntesis, cuando decimos que la energía cinética media de esta partícula es $\frac{3}{2}kT$, sostenemos haber deducido este resultado de la teoría cinética, o sea de las leyes de Newton. Encontraremos que se puede deducir toda clase de cosas —cosas maravillosas— de la teoría cinética, y lo más interesante es el que aparentemente podamos obtener tanto a partir de tan poco. Naturalmente no queremos decir que las leyes de Newton son "poco" —éstas son realmente suficientes—; lo que queremos decir es que *nosotros* no hicimos gran cosa. ¿Y cómo obtenemos tanto? La respuesta es que hemos estado continuamente haciendo una suposición importante: si un sistema determinado está en equilibrio térmico a cierta temperatura, también estará en equilibrio térmico con *cualquier otra cosa* a la misma temperatura. Por ejemplo, si queremos ver cómo se mueve una partícula si estuviera realmente chocando con agua, podemos imaginar que hay un gas compuesto de otra clase de partículas, municiones muy pequeñas que (suponemos) no interactúan con el agua, sino que sólo golpean la partícula con choques "violentos". Si ahora la partícula tiene una saliente; todo lo que nuestras municiones tienen que hacer es golpear la saliente. Sabemos todo lo que hay que saber de este gas imaginario de municiones a temperatura T —es un gas ideal—. El agua es complicada, pero un gas ideal es simple. Ahora bien, *nuestra partícula tiene que estar en equilibrio con el gas de municiones.* En consecuencia, el movimiento medio de la partícula debe ser el que se obtiene de las colisiones en un gas, porque si no se estuviera moviendo a la velocidad adecuada respecto al

agua, sino que, digamos, se estuviera moviendo más rápidamente, eso significaría que las municiones tomarían energía de ella y se pondrían más calientes que el agua. Pero hemos comenzado con ellas a la misma temperatura y suponemos que si algo está una vez en equilibrio, se queda en equilibrio —espontáneamente, no hay partes que se calienten y partes que se enfríen—.

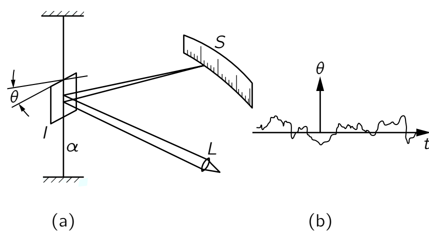


Figura 41-1. (a) Un galvanómetro sensible de espejo. La luz proveniente de una fuente L se refleja en un pequeño espejo y va a una escala. (b) Registro esquemático de la lectura de la escala en función del tiempo.

Esta proposición es cierta y se puede probar a partir de las leyes de la mecánica, pero la demostración es muy complicada y se requiere usar mecánica avanzada. Es mucho más fácil de probar en la mecánica cuántica que en la clásica. Boltzmann fue el primero en demostrarla, pero por ahora pueden simplemente darla por cierta; podemos entonces sostener que nuestra partícula debe tener $\frac{3}{2}kT$ de energía; esto pasa si las municiones artificiales la golpean, por lo que debe haber $\frac{3}{2}kT$ cuando la golpea el agua a la misma temperatura y suprimimos las municiones, por lo tanto es $\frac{3}{2}kT$. Es un razonamiento extraño, pero perfectamente válido.

Además del movimiento de partículas coloidales que dio origen al descubrimiento del movimiento browniano, hay una cantidad de fenómenos, en el laboratorio y en otras situaciones, donde podemos ver movimiento browniano. Si estamos tratando de construir el equipo más delicado posible, digamos que un espejo pequeñísimo sobre una delgada fibra de cuarzo para un galvanómetro balístico muy sensible (figura 41-1), el espejo no se queda quieto, sino que se sacude todo el tiempo —*todo el tiempo*— de modo que cuando lo iluminamos y buscamos la posición de la mancha de luz no tenemos un instrumento perfecto porque el espejo siempre se está sacudiendo. ¿Por qué? Porque la energía cinética media de rotación de este espejo tiene que ser, en promedio, $\frac{1}{2}kT$.

¿Cuál es el ángulo medio cuadrático dentro del cual se balancea el espejo? Supongan que hallamos el periodo natural de vibración del espejo golpeándolo levemente en un lado y viendo cuánto tarda en oscilar ida y vuelta, y que también conocemos el momento de inercia I. Conocemos la expresión de la energía cinética de rotación está dada por la ecuación (19.8): $T = \frac{1}{2}I\omega^2$. Ésta es la energía cinética; la energía potencial que la acompaña será proporcional al cuadrado del ángulo: es $V = \frac{1}{2}\alpha\theta^2$. Pero si conocemos el periodo t_0 y calculamos a partir de él la frecuencia natural $\omega_0 = 2\pi/t_0$, la energía potencial es $V = \frac{1}{2}I\omega_0^2\theta^2$. Ahora bien, sabemos que la energía cinética media es $\frac{1}{2}kT$, pero como es un oscilador armónico la energía potencial media también es $\frac{1}{2}kT$. Luego,

$$\tfrac{1}{2}I\omega_0^2\langle\theta^2\rangle = \tfrac{1}{2}kT,$$

o sea

$$\langle\theta^2\rangle = kT/I\omega_0^2. \tag{41.1}$$

De este modo podemos calcular las oscilaciones del espejo de un galvanómetro y a partir de ello encontrar cuáles serán las limitaciones de nuestro instrumento. Si queremos tener oscilaciones más pequeñas, tenemos que enfriar el espejo. Una cuestión interesante es *dónde* enfriarlo. Depende de dónde está recibiendo los "empujones". Si es a través de la fibra, lo enfriamos en la parte de arriba —si el espejo está rodeado por un gas y está siendo golpeado la mayor parte de las veces por colisiones en el gas, lo mejor es enfriar el gas—. En realidad, si sabemos de dónde viene el *amortiguamiento* de las oscilaciones, ocurre que también ésa es siempre la *fuente* de las fluctuaciones; éste es un punto sobre el cual volveremos más adelante.

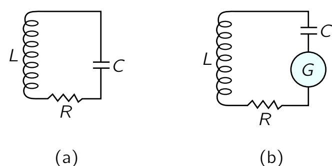


Figura 41-2. Circuito resonante de alto Q. (a) Circuito real a temperatura T. (b) Circuito ficticio con una resistencia ideal (sin ruido) y un "generador de ruido" G.

Aunque parezca sorprendente, pasa lo mismo en los *circuitos eléctricos*. Supongan que estamos haciendo un amplificador muy sensible y preciso para una frecuencia definida y tenemos un circuito resonante (figura 41-2) a la entrada para hacerlo muy sensible a esta frecuencia determinada, como un radiorreceptor, pero uno verdaderamente bueno. Supongan que queremos llegar al límite más bajo posible, así que tomemos el voltaje, digamos de la inductancia, y lo mandamos al resto del amplificador. Naturalmente que en cualquier circuito como éste hay pérdidas. No es un circuito resonante perfecto, es sólo muy bueno y hay, digamos, una pequeña resistencia (introducimos el resistencia para poder verla, pero se supone que es muy pequeña). Ahora nos gustaría saber: ¿cuánto fluctúa la dife-

rencia de potencial entre los extremos de la inductancia? *Respuesta*: sabemos que $\frac{1}{2}LI^2$ es la "energía cinética" —la energía asociada con una bobina en un circuito resonante (capítulo 25)—. En consecuencia, el valor medio de $\frac{1}{2}LI^2$ es igual a $\frac{1}{2}kT$. Esto nos dice cuál es el valor medio cuadrático de la corriente y podemos hallar cuál es el voltaje correspondiente a partir del valor medio cuadrático de la corriente. Puesto que si queremos la diferencia de potencial entre los extremos de la inductancia, la fórmula es $\hat{V}_L = i\omega L\hat{I}$, y el valor medio del módulo al cuadrado de la diferencia de potencial en la inductancia es $\langle V_L^2\rangle = L^2\omega_0^2\langle I^2\rangle$; insertando este valor en $\frac{1}{2}L\langle I^2\rangle = \frac{1}{2}kT$, obtenemos:

$$\langle V_L^2\rangle = L\omega_0^2 kT. \tag{41.2}$$

Por lo tanto, ahora podemos diseñar circuitos y decir cuándo vamos a tener lo que se llama *ruido de Johnson*, ¡el ruido asociado con las fluctuaciones térmicas!

¿Y ahora de dónde provienen las fluctuaciones? Vienen del *resistencia* —provienen del hecho de que los electrones del resistor andan zigzagueando porque están en equilibrio térmico con la materia del resistor, y provocan fluctuaciones en la densidad de electrones—. Producen entonces pequeñísimos campos eléctricos que excitan el circuito resonante.

Los ingenieros eléctricos representan la respuesta de otra manera. Desde el punto de vista físico, el resistor es efectivamente la fuente de ruido. Sin embargo, podemos remplazar el circuito real, que tiene un verdadero resistor físico, por un circuito artificial que contiene un pequeño generador que *representará* el ruido, por lo que ahora el resistor es ideal en este aspecto: ningún ruido proviene de él. Todo el ruido está en el generador artificial. Y así, si conociéramos las características del ruido generado por un resistor, si tuviésemos la fórmula, podríamos calcular qué es lo que el circuito va a hacer en respuesta a ese ruido. Luego, necesitamos una fórmula para las fluctuaciones de ruido. Ahora bien, el ruido generado por el resistor contiene todas las frecuencias, ya que el resistor no es de por sí resonante. Es claro que el circuito resonante sólo "escucha" la parte que está cerca de la frecuencia justa, pero el resistor contiene muchas frecuencias diferentes. Podemos describir cuál es la intensidad del generador de la manera siguiente: la potencia media que absorbería el resistor si estuviera conectado directamente al generador de ruido sería $\langle E^2\rangle/R$, si E fuese el voltaje del generador. Pero querríamos conocer con más detalles cuánta potencia hay en cada frecuencia. En una sola frecuencia hay muy poca potencia; se trata de una distribución. Sea $P(\omega)\,d\omega$ la potencia que entregaría el generador a ese mismo resistor en el intervalo de frecuencia $d\omega$. Podemos demostrar (lo demostraremos para otro caso, pero la matemática es exactamente la misma) que la potencia resulta

$$P(\omega)\,d\omega = (2/\pi)kT\,d\omega, \tag{41.3}$$

y es *independiente de la resistencia* si se expresa de este modo.

41-2 Equilibrio térmico de la radiación

Ahora vamos a considerar un problema aún más avanzado e interesante, que es el siguiente. Supongan que tenemos un oscilador cargado, como los que mencionamos al estudiar la luz, por ejemplo, un electrón oscilando de aquí para allá en un átomo. Si oscila de aquí para allá, irradia luz. Supongamos ahora que este oscilador está en un gas de átomos muy enrarecido y que de vez en cuando los átomos chocan con él. En el equilibrio, después de un largo tiempo, este oscilador ganará energía, de modo que su energía cinética de oscilación será $\frac{1}{2}kT$, y como es un oscilador armónico, toda su energía será kT. Ésta es, naturalmente, una descripción incorrecta hasta aquí, porque el oscilador tiene *carga eléctrica* y si tiene energía kT se está sacudiendo de un lado para otro y *radiando luz*. En consecuencia, es imposible tener equilibrio de la materia real sola, sin que las cargas contenidas en ella emitan luz, y a medida que se emite la luz, la energía se pierde, el oscilador pierde su kT, y así todo el gas que está chocando con el oscilador se enfría gradualmente. Y ésta es, naturalmente, la manera en que una

estufa muy caliente se enfría, radiando la luz hacia el cielo, porque los átomos están sacudiendo su carga y radian continuamente, y poco a poco, a causa de esta radiación, el movimiento disminuye.

Por otra parte, si encerramos todo dentro de una caja de modo que la luz no se vaya al infinito, *podemos* finalmente obtener equilibrio térmico. Podríamos poner el gas en una caja en cuyas paredes podemos decir que hay otros radiadores devolviendo la luz o, para dar un ejemplo más lindo, podemos decir que las paredes de la caja son espejos. Es más fácil pensar en este caso. Suponemos entonces que toda la radiación que sale del oscilador se queda moviendo dentro de la caja. Luego, naturalmente, es cierto que el oscilador comienza a radiar, pero muy pronto puede conservar su kT de energía a pesar de que está radiando, porque está iluminado, por así decir, con su propia luz reflejada en las paredes de la caja. Es decir, que después de un rato hay una gran cantidad de luz aglomerándose en la caja, y aunque el oscilador radie un poco, la luz vuelve y restituye parte de la energía radiada.

Determinaremos ahora cuánta luz debe haber en esa caja a la temperatura T para que la luz que cae sobre este oscilador genere precisamente la energía necesaria para compensar la luz que ha irradiado.

Supongamos que los átomos del gas sean muy pocos y estén muy separados, de modo que tengamos un oscilador ideal sin resistencia con excepción de la resistencia de radiación. Luego consideramos que en el equilibrio térmico el oscilador está haciendo dos cosas al mismo tiempo. Primero: tiene una energía media kT y podemos calcular cuánta radiación emite. Segundo: esta radiación debe ser exactamente la cantidad que resultaría del hecho de que la luz que incide sobre el oscilador se dispersa. Como no hay ningún otro lado a donde pueda ir la energía, esta radiación efectiva es luz dispersada proveniente de la luz presente.

Por lo tanto, calculemos primero la energía radiada por el oscilador en un segundo, si el oscilador tiene cierta energía. (Tomamos unas cuantas ecuaciones del capítulo 32 sobre la resistencia de radiación sin volver a derivarlas.) La energía radiada por radián dividida entre la energía del oscilador se denomina $1/Q$ (ecuación 32.8): $1/Q = (dW/dt)/\omega_0 W$. Usando la constante de amortiguamiento γ, también se puede escribir $1/Q = \gamma/\omega_0$, donde ω_0 es la frecuencia natural del oscilador —si γ es muy pequeña, Q es muy grande—. La energía radiada por segundo es entonces

$$\frac{dW}{dt} = \frac{\omega_0 W}{Q} = \frac{\omega_0 W \gamma}{\omega_0} = \gamma W. \tag{41.4}$$

La energía radiada por segundo es, pues, simplemente gamma por la energía del oscilador. Ahora bien, el oscilador debería tener una energía media kT, por lo que vemos que gamma kT es la cantidad media de energía radiada por segundo:

$$\langle dW/dt \rangle = \gamma kT. \tag{41.5}$$

Ahora sólo tenemos que saber qué es gamma. Se puede encontrar gamma fácilmente con la ecuación (32.12). Es

$$\gamma = \frac{\omega_0}{Q} = \frac{2}{3}\,\frac{r_0 \omega_0^2}{c}, \tag{41.6}$$

donde $r_0 = e^2/mc^2$ es el radio clásico del electrón, y hemos puesto $\lambda = 2\pi c/\omega_0$.

Nuestro resultado final para la radiación media de luz por unidad de tiempo cerca de la frecuencia ω_0 es entonces

$$\langle dW/dt \rangle = \frac{2}{3}\,\frac{r_0 \omega_0^2 kT}{c}. \tag{41.7}$$

Ahora preguntamos cuánta luz debe estar incidiendo sobre el oscilador. Debe bastar para que la energía absorbida de la luz (y después dispersada) sea exactamente esa cantidad. En otras palabras, la luz emitida se explica como luz *dispersada* proveniente de la luz que incide sobre el oscilador en la cavidad. Por lo tanto, tenemos que calcular ahora cuánta luz se dispersa en el oscilador si hay

cierta cantidad —desconocida— de radiación incidiendo sobre él. Sea $I(\omega)d\omega$ la cantidad de energía luminosa que hay a la frecuencia ω, dentro de un intervalo $d\omega$ (porque no hay luz a una frecuencia *exacta*, está distribuida por todo el espectro). Así, $I(\omega)$ es cierta *distribución espectral* que ahora vamos a hallar —es el color de un horno a temperatura T que vemos al abrir la puerta y mirar por el agujero—. Ahora, ¿cuánta luz se absorbe? Ya calculamos la cantidad de radiación absorbida de un haz dado de luz incidente y lo hicimos en términos de una *sección eficaz*. Es justamente como si dijéramos que se absorbe toda la luz que incide sobre cierta sección transversal. Así, pues, la cantidad total que se vuelve a radiar (se dispersa) es la intensidad incidente $I(\omega)\,d\omega$ multiplicada por la sección eficaz σ.

La fórmula que obtuvimos para la sección eficaz, ecuación (32.19), no contenía el amortiguamiento. No es difícil repetir la derivación incluyendo el término de resistencia que despreciamos. Si lo hacemos y calculamos la sección eficaz en la misma forma obtenemos

$$\sigma_s = \frac{8\pi r_0^2}{3}\left(\frac{\omega^4}{(\omega^2-\omega_0^2)^2+\gamma^2\omega^2}\right). \tag{41.8}$$

Ahora bien, en función de la frecuencia, σ_s tiene un valor apreciable sólo para ω cerca de la frecuencia natural ω_0. (Recuerden que el Q de un oscilador radiante es del orden de 10^8.) El oscilador dispersa fuertemente cuando ω es igual a ω_0 y muy débilmente para otros valores de ω. Por lo tanto, podemos remplazar ω por ω_0 y $\omega^2-\omega_0^2$ por $2\omega_0(\omega-\omega_0)$, obteniendo:

$$\sigma_s = \frac{2\pi r_0^2\omega_0^2}{3[(\omega-\omega_0)^2+\gamma^2/4]}. \tag{41.9}$$

Ahora toda la curva está localizada cerca de $\omega=\omega_0$. (En realidad no tenemos que hacer ninguna aproximación, pero es mucho más fácil hacer las integrales si simplificamos un poco las ecuaciones.) Ahora multiplicamos la intensidad en un intervalo dado de frecuencia por la sección eficaz de dispersión para obtener la cantidad de energía dispersada en el intervalo $d\omega$. La energía *total* dispersada es entonces la integral sobre todo ω. Luego,

$$\begin{aligned}\frac{dW_s}{dt} &= \int_0^\infty I(\omega)\sigma_s(\omega)\,d\omega \\ &= \int_0^\infty \frac{2\pi r_0^2\omega_0^2 I(\omega)\,d\omega}{3[(\omega-\omega_0)^2+\gamma^2/4]}.\end{aligned} \tag{41.10}$$

Hagamos ahora $dW_s/dt = 3\gamma kT$. ¿Por qué *tres*? Porque cuando hicimos el análisis de la sección eficaz en el capítulo 32 supusimos que la polarización era tal que la luz podía forzar el oscilador. Si hubiéramos usado un oscilador que sólo se pudiera mover en una dirección y la luz estuviera, por ejemplo, polarizada en dirección errada, no podría dar dispersión. Por lo tanto, o debemos promediar la sección eficaz de un oscilador que sólo puede ir en una dirección sobre todas las direcciones de incidencia y de polarización de la luz, o, lo que es más fácil, podemos imaginar un oscilador que seguirá el campo, cualquiera que sea la dirección en que apunte el campo. Este oscilador, que puede oscilar de igual modo en tres direcciones, tendría $3kT$ de energía media porque hay tres grados de libertad en el mismo. Es porque hay tres grados de libertad que usamos $3\gamma kT$.

Ahora tenemos que hacer la integral. Supongamos que la distribución espectral desconocida de la luz, $I(\omega)$, es una curva suave y que no varía mucho en la estrechísima región de frecuencia donde σ_s tiene el pico (figura 41-3). Entonces la única contribución apreciable proviene de cuando ω está muy cerca de ω_0, dentro de una cantidad gamma que es muy pequeña. En consecuencia, aunque $I(\omega)$ sea una función desconocida y complicada, el único lugar donde es importante es cerca de $\omega=\omega_0$ y allí podemos remplazar la curva suave por una horizontal —una constante— a la misma altura. En otras palabras, sacamos $I(\omega)$ fuera de la integral y la llamamos $I(\omega_0)$. También podemos sacar el resto de las constantes fuera de la integral y lo que nos queda es

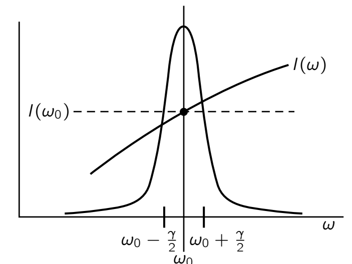


Figura 41-3. Los factores del integrando de (41.10). El pico es la curva de resonancia $1/[(\omega-\omega_0)^2+\gamma^2/4]$. Con buena aproximación se puede remplazar el factor $I(\omega)$ por $I(\omega_0)$.

$$\tfrac{2}{3}\pi r_0^2\omega_0^2 I(\omega_0)\int_0^\infty \frac{d\omega}{(\omega-\omega_0)^2+\gamma^2/4} = 3\gamma kT. \tag{41.11}$$

Ahora bien, la integral debería ir de 0 a ∞, pero el 0 está tan lejos de ω_0 que para entonces la curva ya se ha acabado, por lo que en su lugar iremos hasta menos ∞ —da prácticamente lo mismo y es mucho más fácil hacer la integral—. La integral es una función arco tangente de la forma $\int dx/(x^2+a^2)$. Si la buscamos en un libro vemos que es igual a π/a. Por lo tanto, en nuestro caso resulta $2\pi/\gamma$. En consecuencia, reordenando un poco obtenemos:

$$I(\omega_0) = \frac{9\gamma^2 kT}{4\pi^2 r_0^2\omega_0^2}. \tag{41.12}$$

Luego sustituimos la expresión (41.6) de gamma (no se molesten en escribir ω_0; como es válida para cualquier ω_0, la podemos llamar simplemente ω) y la fórmula que resulta para $I(\omega)$ es:

$$I(\omega) = \frac{\omega^2 kT}{\pi^2 c^2}. \tag{41.13}$$

Y ésta es la distribución de luz en un horno muy caliente. Se llama *radiación de cuerpo negro*. Negro, porque cuando la temperatura es cero el agujero del horno al cual estamos mirando es negro.

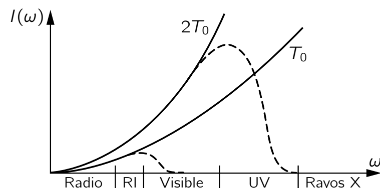


Figura 41-4. Distribución de intensidad de la radiación de cuerpo negro a dos temperaturas, según la física clásica (curvas llenas). Las curvas discontinuas muestran la distribución real.

Según la teoría clásica, la ecuación (41.13) es la distribución de energía de radiación dentro de una caja cerrada a temperatura T. En primer lugar, notemos una característica notable de esa expresión. La carga del oscilador, la masa del oscilador, todas las propiedades específicas del oscilador, *se cancelan* porque, una vez alcanzado el equilibrio con un oscilador debe haber equilibrio con cualquier otro oscilador de masa diferente, o nos encontraremos en dificultades. Así, pues, éste es un tipo importante de verificación del aserto; el equilibrio no depende de con qué estemos en equilibrio, sino *únicamente de la temperatura*. Dibujemos ahora la curva de $I(\omega)$ (figura 41-4). Nos dice cuánta luz tenemos para cada frecuencia.

La intensidad contenida en nuestra caja por unidad de intervalo de frecuencia varía, como vemos, con el cuadrado de la frecuencia, lo cual significa que si tenemos una caja a cualquier temperatura y miramos los rayos X que salen, ¡habrá una cantidad!

Naturalmente, sabemos que esto es falso. Cuando abrimos el horno y le damos una ojeada no nos quemamos los ojos con los rayos X. Es completamente falso. Aún más, la *energía total* que hay dentro de la caja, el total de toda esta intensidad sumada sobre todas las frecuencias, sería el área debajo de esta curva infinita. Por lo tanto, está fundamental, poderosa y absolutamente errado.

Es así como la teoría clásica fue *absolutamente incapaz* de describir correctamente la distribución de la luz proveniente de un cuerpo negro, como fue incapaz de describir correctamente los calores específicos de los gases. Los físicos volvieron una y otra vez sobre esta derivación desde muchos puntos de vista y no había escapatoria: *es* la predicción de la física clásica. La ecuación (41.13) se llama *ley de Rayleigh*, es la predicción de la física clásica y es evidentemente absurda.

41-3 La equipartición y el oscilador cuántico

La dificultad anterior es otro aspecto del prolongado problema de la física clásica que comenzó con la dificultad de los calores específicos de los gases y que ahora ha sido enfocado sobre la distribución de la luz en un cuerpo negro. Ahora bien, como es natural, en la época en que los teóricos estudiaron esto había también muchas *medidas* de la curva verdadera. Y ocurría que la curva correcta era como las curvas discontinuas de la figura 41-4. Es decir que no había rayos X. Si bajamos la temperatura, toda la curva se baja proporcionalmente a T, de acuerdo con la teoría clásica, pero la curva observada también se corta más pronto a baja temperatura. Por lo tanto, el extremo de baja frecuencia de la curva es

correcto, pero el extremo de alta frecuencia está mal. ¿Por qué? Mientras sir James Jeans se preocupaba de los calores específicos de los gases, notó que los movimientos que tienen frecuencia alta se "congelan" cuando la temperatura se hace muy baja. O sea, que si la temperatura es muy baja, si la frecuencia es muy alta, los osciladores *no tienen* kT de energía en promedio. Ahora bien, recordemos cómo funcionaba nuestra deducción de (41.13), todo depende de la energía de un oscilador en equilibrio térmico. Lo que el kT de (41.5) era y lo que es el mismo kT en (41.13), es la energía media de un oscilador armónico de frecuencia ω a temperatura T. Clásicamente es kT pero ¡experimentalmente no! —no cuando la temperatura es demasiado baja o la frecuencia del oscilador es demasiado alta—. Así pues, la razón de que la curva caiga es la misma razón de que los calores específicos de los gases estén errados. Es más fácil estudiar la curva del cuerpo negro que los calores específicos de los gases, que son tan complicados, por lo que enfocaremos nuestra atención en determinar la verdadera curva del cuerpo negro, porque nos dice correctamente, a cada frecuencia, cuál es verdaderamente la energía media de un oscilador armónico en función de la temperatura.

Planck estudió esta curva. Primero halló la respuesta empíricamente ajustando la curva observada con una linda función que andaba muy bien. Así, tenía una fórmula empírica para la energía media de un oscilador en función de la frecuencia. En otras palabras: tenía la fórmula *correcta* en lugar de kT y después de jugar un poco encontró una derivación simple basada en una hipótesis muy extraña. Esa hipótesis era que el *oscilador armónico sólo puede adquirir energía de* $\hbar\omega$ *por vez*. La idea de que puedan tener *absolutamente cualquier energía* es falsa. Por supuesto que aquello fue el principio del fin de la mecánica clásica.

Derivaremos ahora la primerísima fórmula cuántica determinada correctamente. Supongan que los niveles de energía permitidos en un oscilador armónico estén igualmente espaciados a una distancia $\hbar\omega_0$ de modo que el oscilador sólo pueda tomar estas diversas energías (figura 41-5). Planck hizo un razonamiento algo más complicado que el que vamos a dar aquí, ya que era el mismísimo comienzo de la mecánica cuántica y tuvo que demostrar algunas cosas. Pero nosotros tomaremos como un hecho (que en este caso él demostró) que la probabilidad de ocupar un nivel de energía E es $P(E) = \alpha e^{-E/kT}$. Si procedemos así, obtendremos el resultado correcto.

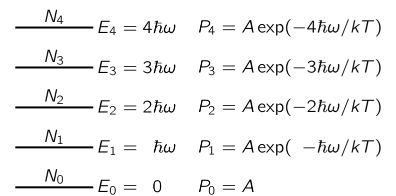


Figura 41-5. Los niveles de energía de un oscilador armónico están igualmente espaciados: $E_n = n\hbar\omega$.

Supongan ahora que tenemos muchos osciladores, siendo cada uno un vibrador de frecuencia ω_0. Algunos de estos vibradores estarán en el estado cuántico más bajo, otros en el siguiente y así sucesivamente. Lo que querríamos conocer es la energía media de todos estos osciladores. Para encontrarla calculemos la energía total de todos los osciladores y dividámosla entre el número de osciladores. Ésa será la energía media por oscilador en el equilibrio térmico, y también será la energía que está en equilibrio con la radiación de cuerpo negro y que debe ir en la ecuación (41.13) en vez de kT. Así, sea N_0 el número de osciladores que están en el estado fundamental (el estado de energía más baja); N_1 el número de osciladores en el estado E_1; N_2 el número de los que están en E_2, etc. Conforme a la hipótesis (que no hemos demostrado) de que en la mecánica cuántica la ley que remplaza la probabilidad $e^{-\text{EP}/kT}$ o $e^{-\text{EC}/kT}$ de la mecánica clásica, es que la probabilidad decrece como $e^{-\Delta E/kT}$, donde ΔE es la energía en exceso, supondremos que el número N_1 de los que están en el primer estado será el número N_0 de los que están en el estado fundamental, multiplicado por $e^{-\hbar\omega/kT}$. Análogamente, el número N_2 de osciladores en el segundo estado es $N_2 = N_0 e^{-2\hbar\omega/kT}$. Para simplificar el álgebra, pongamos $e^{-\hbar\omega/kT} = x$. Entonces tendremos simplemente $N_1 = N_0 x$, $N_2 = N_0 x^2$, $\ldots$, $N_n = N_0 x^n$.

Primero se debe calcular la energía total de todos los osciladores. Si un oscilador está en el estado fundamental, no hay energía. Si está en el primer estado, la energía es $\hbar\omega$ y hay N_1 de ellos. Luego, $N_1\hbar\omega$, o $\hbar\omega N_0 x$, es la cantidad de energía que obtenemos de ellos. Los que están en el segundo estado tienen $2\hbar\omega$ y son N_2 en total, por lo que $N_2 \cdot 2\hbar\omega = 2\hbar\omega N_0 x^2$ es la cantidad de energía que obtenemos, y así sucesivamente. Luego los sumamos todos obteniendo $E_{\text{tot}} = N_0\hbar\omega(0 + x + 2x^2 + 3x^3 + \ldots)$.

Y ahora, ¿cuántos osciladores hay? Naturalmente, N_0 es el número de los que están en el estado fundamental, N_1 en el primer estado, etc., y los sumamos: $N_{\text{tot}} = N_0(1 + x + x^2 + x^3 + \dots)$. La energía media, entonces es

$$\langle E\rangle = \frac{E_{\text{tot}}}{N_{\text{tot}}} = \frac{N_0\hbar\omega(0 + x + 2x^2 + 3x^3 + \cdots)}{N_0(1 + x + x^2 + x^3 + \cdots)}. \tag{41.14}$$

Dejamos al lector las dos sumas que aparecen aquí para que jueguen y se diviertan. Una vez terminadas las sumas, remplazando x en el resultado, debemos obtener —si no cometemos errores en la suma—

$$\langle E\rangle = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \tag{41.15}$$

Ésta fue, pues, la primera fórmula cuántica que se conoció, o se discutió, y fue la magnífica culminación de décadas de rompederos de cabeza. Maxwell sabía que algo estaba mal y el problema era ¿qué era lo que estaba *bien*? Aquí está la respuesta cuantitativa de lo que está bien en vez de kT. Naturalmente, esta expresión se debe aproximar a kT cuando $\omega \to 0$ o cuando $T \to \infty$. Vean si pueden demostrar que es así —aprendan a usar la matemática—.

Éste es el famoso factor de corte que Jeans estaba buscando, y si lo usamos en lugar de kT en (41.13) obtenemos, para la distribución de la luz en una caja negra:

$$I(\omega)\,d\omega = \frac{\hbar\omega^3\,d\omega}{\pi^2c^2(e^{\hbar\omega/kT} - 1)}. \tag{41.16}$$

Vemos que para ω grande, aunque tengamos ω^3 en el numerador hay una e elevada a una potencia enorme en el denominador, por lo que la curva vuelve a bajar y no "salta por el aire" —¡no obtenemos ni luz ultravioleta ni rayos X donde no los esperamos!—.

Alguien podría quejarse de que en nuestra derivación de la ecuación (41.16) empleamos la teoría cuántica para los niveles de energía del oscilador y la teoría clásica en la determinación de la sección eficaz σ_s. Pero la teoría cuántica de la luz en interacción con un oscilador armónico da exactamente el mismo resultado que la teoría clásica. En realidad es por eso que se justificaba gastar tanto tiempo en nuestro análisis del índice de refracción y de la dispersión de la luz utilizando un modelo de átomos como pequeños osciladores: las fórmulas cuánticas son sustancialmente las mismas.

Volvamos ahora al ruido de Johnson en un resistor. Ya hemos señalado que la teoría de esta potencia de ruido es en realidad la misma que la teoría de la distribución clásica de cuerpo negro. En verdad, de manera bastante sorprendente; ya hemos dicho que si la resistencia de un circuito no fuera una resistencia real, sino una antena (una antena actúa como una resistencia porque radia energía), una resistencia de radiación, nos resultaría fácil calcular cuál sería la potencia. Sería justamente la potencia que entra en la antena proveniente de la luz que la rodea, y obtendríamos la misma distribución, a excepción de uno o dos factores. Podemos suponer que el resistor es un generador con un espectro desconocido $P(\omega)$ de potencia. El espectro está determinado por el hecho de que este mismo generador, conectado a un circuito resonante de *cualquier frecuencia*, como en la figura 41-2(b), genera en la inductancia un voltaje dado por la ecuación (41.2). Se llega entonces a la misma integral que en (41.10) y el mismo método da la ecuación (41.3). Para temperaturas bajas, el kT de (41.3) se debe remplazar, naturalmente, por (41.15). Las dos teorías (radiación de cuerpo negro y ruido de Johnson) también están íntimamente relacionadas desde el punto de vista físico, ya que, por supuesto, podemos conectar un circuito resonante a una *antena*, por lo que la resistencia R será exclusivamente una *resistencia de radiación.* Como (41.2) no depende del origen físico de la resistencia, sabemos que el generador G para una resistencia real es el mismo que para una resistencia de radiación. ¿Cuál es el origen de la potencia $P(\omega)$ generada si la resistencia R es sólo una antena ideal en equilibrio con el medio a temperatura T? Es la radiación $I(\omega)$ en

el espacio a temperatura T que incide sobre la antena y, como "señales recibidas", hace un generador efectivo. Por lo tanto, se puede deducir una relación directa entre $P(\omega)$ e $I(\omega)$, que *lleva* entonces de (41.13) a (41.3).

Todas las cosas de que hemos estado hablando —el llamado ruido de Johnson y la distribución de Planck, y la teoría correcta del movimiento browniano que estamos por describir— son desarrollos de la primera década del siglo XX, más o menos. Ahora, con estos puntos y esta historia en mente, volvamos al movimiento browniano.

41-4 La caminata aleatoria

Consideremos cómo debe variar en el tiempo la posición de una partícula que se agita, para tiempos muy largos respecto al tiempo entre "golpes". Consideren una pequeña partícula en movimiento browniano que está zigzagueando porque las moléculas de agua, en zigzagueo irregular, la bombardean de todos lados. Pregunta: después de un intervalo dado de tiempo, ¿a qué distancia de donde salió es probable que se encuentre? Este problema fue resuelto por Einstein y Smoluchowski. Si imaginamos dividir el tiempo en pequeños intervalos, digamos que de un centésimo de segundo más o menos, después del primer centésimo de segundo se mueve hasta aquí, en el siguiente centésimo se mueve un poco más, en el siguiente centésimo se mueve para otro lado, y así sucesivamente. Frente a la rapidez del bombardeo, un centésimo de segundo es un tiempo muy largo. El lector puede verificar fácilmente que el número de colisiones que una sola molécula de agua recibe en un segundo es del orden de 10^{14} por lo que en un centésimo de segundo sufre 10^{12} colisiones, ¡que es bastante! Por lo tanto, después de un centésimo de segundo no recordará lo que le pasó antes. En otras palabras, las colisiones son todas *al azar*, de modo que un "paso" no está relacionado con el "paso" precedente. Es como el famoso problema del marinero borracho: el marinero sale del bar y da una serie de pasos, pero da cada uno a un ángulo arbitrario, al azar (figura 41-6). La pregunta es: después de mucho tiempo, ¿dónde está el marinero? Naturalmente, ¡no lo sabemos! Es imposible decirlo. ¿Qué entendemos por esto que está en alguna parte más o menos al azar? Bueno, ¿dónde está el promedio entonces? *En promedio, ¿a qué distancia del bar ha llegado?* Ya hemos contestado esta pregunta, porque una vez estudiamos la superposición de la luz proveniente de muchas fuentes diferentes con fases distintas y eso significaba sumar una gran cantidad de flechas a diversos ángulos (capítulo 32). Allí descubrimos que el promedio del cuadrado de la distancia entre un extremo y otro de la cadena de pasos al azar, que era la intensidad de la luz, era la suma de las intensidades de los pedazos separados. Así, pues, con el mismo tipo de matemática, podemos demostrar inmediatamente que si $\boldsymbol{R}_N$ es el vector distancia al origen después de N pasos, el promedio del cuadrado de la distancia al origen es proporcional al número N de pasos. Es decir, $\langle R_N^2 \rangle = NL^2$, donde L es la longitud de cada paso. Como el número de pasos es proporcional al tiempo en nuestro problema actual, *el valor del cuadrado de la distancia es proporcional al tiempo*:

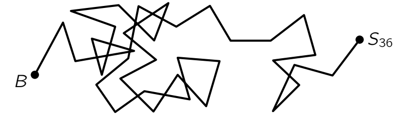


Figura 41-6. caminata aleatoria de 36 pasos de longitud l. ¿A qué distancia está S_{36} de B? *Respuesta*: alrededor de $6l$ en promedio.

$$\langle R^2 \rangle = \alpha t. \tag{41.17}$$

Esto no significa que la *distancia media* sea proporcional al tiempo. Si la distancia media fuera proporcional al tiempo, eso significaría que el movimiento neto se realiza a una buena velocidad uniforme. El marinero *está* avanzando de manera relativamente apreciable, pero únicamente de modo que el promedio del *cuadrado* de la distancia es proporcional al tiempo. Ésta es la característica de una caminata aleatoria.

Podemos demostrar muy fácilmente que en cada paso sucesivo el cuadrado de la distancia aumenta L^2 en promedio. Puesto que si escribimos $\boldsymbol{R}_N = \boldsymbol{R}_{N-1} + \boldsymbol{L}$, encontramos que $\boldsymbol{R}_N^2$ es

$$\boldsymbol{R}_N \cdot \boldsymbol{R}_N = R_N^2 = R_{N-1}^2 + 2\boldsymbol{R}_{N-1} \cdot \boldsymbol{L} + L^2,$$

y promediando sobre muchos intentos, tenemos $\langle R_N^2 \rangle = \langle R_{N-1}^2 \rangle + L^2$, ya que $\langle \boldsymbol{R}_{N-1} \cdot \boldsymbol{L} \rangle = 0$. Luego, por inducción,

$$\langle R_N^2 \rangle = NL^2. \tag{41.18}$$

Ahora querríamos calcular el coeficiente α que aparece en la ecuación (41.17), y para hacerlo tenemos que agregar algo. Vamos a suponer que si le aplicásemos una fuerza a esta partícula (que no tiene nada que ver con el movimiento browniano; estamos encarando un problema marginal por el momento), la misma reaccionaría contra la fuerza de la siguiente manera. Primeramente, habría inercia. Sea m el coeficiente de inercia, la masa efectiva del objeto (no necesariamente igual a la masa real de la partícula real, porque el agua se tiene que mover alrededor de la partícula si tiramos de ella). Así, si hablamos de movimiento en una dirección hay por un lado un término $m(d^2x/dt^2)$. Y también queremos suponer que si continuáramos tirando regularmente del objeto, habría una resistencia al avance, proveniente del fluido, la cual es proporcional a la velocidad. Además de la inercia del fluido, hay una resistencia al movimiento debido a la viscosidad y a la complejidad del fluido. Es absolutamente esencial que *haya* alguna pérdida irreversible, algo parecido a una resistencia, para que haya fluctuaciones. No hay manera de producir el kT a no ser que también haya pérdidas. La fuente de las fluctuaciones está muy relacionada con estas pérdidas. Dentro de poco discutiremos el mecanismo de esta resistencia al avance —hablaremos de fuerzas proporcionales a la velocidad y de dónde provienen—. Pero por ahora supongamos que esta resistencia existe. Entonces, cuando estamos tirando de la partícula de manera corriente, la fórmula para el movimiento bajo la acción de una fuerza externa es:

$$m\,\frac{d^2x}{dt^2} + \mu\,\frac{dx}{dt} = F_{\text{ext}}. \tag{41.19}$$

La cantidad μ se puede determinar directamente del experimento. Por ejemplo, podemos observar la caída de la gota por gravedad. Entonces sabemos que la fuerza es mg y μ es mg dividido entre la velocidad de caída que adquiere finalmente la gota. O también podríamos poner la gota en una centrífuga y ver cuánto tarda en sedimentar. O bien, si está cargada, le podemos aplicar un campo eléctrico. Por lo tanto, μ es algo medible, no una cosa artificial, y se conoce para muchos tipos de partículas coloidales, etcétera.

Usemos ahora la misma fórmula en el caso en que la fuerza no es externa, sino igual a las fuerzas irregulares del movimiento browniano. Trataremos entonces de determinar el valor medio del cuadrado de la distancia que recorre el objeto. En vez de tomar las distancias en tres dimensiones, tomemos sencillamente una dimensión y hallemos el valor medio de x^2, simplemente para prepararnos. (Evidentemente, el promedio de x^2 es igual al promedio de y^2 que es igual al promedio de z^2, y por lo tanto, el promedio del cuadrado de la distancia es justamente 3 veces lo que vamos a calcular.) La componente x de las fuerzas irregulares es, naturalmente, tan irregular como cualquiera de las otras componentes. ¿Cuál es la rapidez de variación de x^2? Es $d(x^2)/dt = 2x(dx/dt)$, y así, lo que tenemos que hallar es el promedio de la posición por la velocidad. Demostraremos que es una constante y que por lo tanto el promedio del cuadrado del radio aumentará proporcionalmente al tiempo, y veremos con qué rapidez. Ahora bien, si multiplicamos la ecuación (41.19) por x, $mx(d^2x/dt^2) + \mu x(dx/dt) = xF_x$. Queremos el promedio temporal de $x(dx/dt)$, hagamos entonces el promedio de toda la ecuación y estudiemos los tres términos. Ahora bien, ¿qué pasa con x multiplicada por la fuerza? Si ocurre que la partícula ha recorrido cierta distancia x, el siguiente impulso puede estar en cualquier dirección respecto a x, ya que la fuerza irregular es *completamente* irregular y no sabe de dónde partió la partícula. Si x es positiva, no hay razón para que la fuerza media esté también en esa dirección. Es tan probable que esté en una como en otra. Las fuerzas de bombardeo no la están arrastrando en una dirección determinada. Por lo tanto, el valor medio de x por F es cero. Por otro lado, tendremos que usar un poco de imaginación con el término $mx(d^2x/dt^2)$ y escribirlo en la forma:

$$mx\,\frac{d^2x}{dt^2} = m\,\frac{d[x(dx/dt)]}{dt} - m\left(\frac{dx}{dt}\right)^2.$$

Así, introducimos estos dos términos y hacemos el promedio de ambos. Veamos entonces cuánto debe ser el primer término. Ahora, x por la velocidad tiene un promedio que no varía en el tiempo, porque cuando llega a una posición no se acuerda de dónde estaba antes y de ahí que las cosas ya no cambien con el tiempo. Por lo tanto, esta cantidad es, en promedio, cero. Nos queda la cantidad mv^2 y ésta es la única cosa que sabemos: el valor medio de $mv^2/2$ es $\frac{1}{2}kT$. En consecuencia, encontramos que

$$\left\langle mx\,\frac{d^2x}{dt^2}\right\rangle + \mu\left\langle x\,\frac{dx}{dt}\right\rangle = \langle xF_x\rangle$$

implica

$$-\langle mv^2\rangle + \frac{\mu}{2}\,\frac{d}{dt}\,\langle x^2\rangle = 0,$$

o sea

$$\frac{d\langle x^2\rangle}{dt} = 2\,\frac{kT}{\mu}. \tag{41.20}$$

En consecuencia, después de un cierto tiempo t, el objeto tiene una media del cuadrado de la distancia, $\langle R^2\rangle$, igual a

$$\langle R^2\rangle = 6kT\,\frac{t}{\mu}. \tag{41.21}$$

¡Y así podemos determinar realmente *a qué distancia* llegan las partículas! Debemos determinar primero cómo reaccionan a una fuerza constante, a qué velocidad se desplazan bajo la acción de una fuerza conocida (para hallar μ), y entonces podemos determinar hasta dónde llegan en sus movimientos al azar. Esta ecuación tuvo una importancia histórica considerable porque fue una de las primeras maneras de determinar la constante k. Después de todo, podemos medir μ, el tiempo, hasta dónde llegan las partículas, y podemos hacer el promedio. La razón de que la determinación de k fuera importante es que en la ley $PV = RT$ para un mol, sabemos que R, que también se puede medir, es igual al número de átomos que hay en un mol multiplicado por k. El mol fue definido originalmente como tantos *gramos* de oxígeno 16 (ahora se usa el carbono), por lo que originalmente no se conocía el número de *átomos* que hay en un mol. Naturalmente, es un problema muy interesante e importante. ¿Cómo son de grandes los átomos? ¿Cuántos hay? Así, pues, una de las primeras determinaciones del número de átomos fue por medio de la determinación de hasta dónde se mueve una particulita sucia, observándola pacientemente al microscopio durante un cierto intervalo de tiempo. Y así se determinó la constante k de Boltzmann y el número de Avogadro N_0 porque R ya había sido medida.

42

Aplicaciónes de la teoría cinética

42-1 Evaporación

En este capítulo discutiremos otras aplicaciones de la teoría cinética. En el capítulo anterior recalcamos un aspecto en particular de la teoría cinética: la energía cinética promedio correspondiente a cualquier grado de libertad de una molécula u otro objeto es $\frac{1}{2}kT$. Por otro lado, la característica central que discutiremos ahora es que la probabilidad de encontrar una partícula en lugares diferentes, por unidad de volumen, varía como $e^{-\text{energía potencial}/kT}$; haremos un cierto número de aplicaciones de ello.

Los fenómenos que queremos estudiar son relativamente complicados: un líquido que se evapora, electrones que están saliendo de la superficie de un metal o una reacción química en la que interviene un gran número de átomos. En tales casos, ya no es posible hacer cualquier afirmación correcta y simple a partir de la teoría cinética, ya que la situación es demasiado complicada. Por lo tanto, este capítulo no tiene mucha exactitud, excepto donde lo digamos explícitamente. La idea que hay que recalcar es solamente que a partir de la teoría cinética podemos entender *más o menos* cómo deberían comportarse las cosas. Usando razonamientos termodinámicos, o ciertas medidas empíricas de algunas cantidades críticas, podemos obtener una representación más precisa de los fenómenos.

Sin embargo, es muy útil conocer aunque sea poco más o menos por qué algo se comporta como lo hace, y así, cuando nos encontremos con una situación nueva o una que no hayamos comenzado aún a analizar, podemos decir más o menos lo que debe suceder. Por ello, esta discusión es imprecisa en grado sumo, pero es esencialmente buena —buena en la idea, pero digamos que un poquito simplificada en los detalles específicos—.

El primer ejemplo que consideraremos es la evaporación de un líquido. Supongan que tenemos una caja de gran volumen, parcialmente llena con líquido en equilibrio y con el vapor a una cierta temperatura. Supondremos que las moléculas de vapor están relativamente alejadas unas de otras y que las del líquido están amontonadas. El problema es cuantificar el número de moléculas que hay en la fase vapor, en comparación con el número de moléculas del líquido. ¿Cuál es la densidad del vapor a una temperatura dada y cómo depende de ella?

Llamemos n al número de moléculas por unidad de volumen del vapor. Naturalmente, este número varía con la temperatura. Si calentamos, tendremos una evaporación mayor. Digamos también que $1/V_a$ es el número de átomos del líquido por unidad de volumen. Suponemos que cada molécula del líquido ocupa un cierto volumen, de modo que si hay más moléculas de líquido, todas juntas ocupan un volumen mayor. Por lo que si V_a es el volumen ocupado por una molécula, el número de moléculas en una unidad de volumen es una unidad de volumen dividida entre el volumen de cada molécula. Además, suponemos que hay una fuerza de atracción entre las moléculas que las mantiene juntas en el líquido. De otra forma no podemos entender por qué se condensa. Así, suponemos que hay esa fuerza y que hay una energía de ligadura de las moléculas en el líquido que se pierde cuando pasan a vapor. O sea, si se tiene que hacer una cierta cantidad de trabajo W para pasar una sola molécula del líquido al vapor, entonces hay una cierta diferencia W entre la energía de una molécula en el líquido y la que tendría si estuviese en el vapor, ya que hemos tenido que separarla de las otras moléculas que la atraían.

Ahora usamos el principio general de que el número de átomos por unidad de volumen en dos regiones diferentes es $n_2/n_1 = e^{-(\mathrm{E}_2-\mathrm{E}_1)/kT}$. Por lo que el

número n por unidad de volumen en el vapor, dividido entre el número $1/V_a$ por unidad de volumen en el líquido, es igual a

$$nV_a = e^{-W/kT}, \tag{42.1}$$

ya que ésta es la regla general. Es como la atmósfera en equilibrio bajo la acción de la gravedad, donde el gas es más denso abajo que arriba debido al trabajo mgh que se necesita para levantar las moléculas de gas a la altura h. En el líquido, las moléculas son más densas que en el vapor porque tenemos que sacarlas atravesando la “montaña” de energía W, y la razón de las densidades es $e^{-W/kT}$.

Esto es lo que queríamos deducir: que la densidad del vapor varía como e elevado a alguna energía negativa sobre kT. Los factores que se encuentren delante no son realmente interesantes, ya que en la mayoría de los casos la densidad del vapor es mucho más pequeña que la del líquido. Bajo estas circunstancias, en que no estamos cerca del punto crítico donde las dos son casi iguales, sino que la densidad del vapor es mucho más pequeña que la del líquido, el hecho de que n es mucho más pequeño que $1/V_a$ se debe a que W es muchísimo más grande que kT. Por lo tanto, las fórmulas como (42.1) son interesantes solamente cuando W es muchísimo más grande que kT, porque así, como estamos elevando e a una cantidad negativa tremenda, si cambiamos T un poquito, ese enorme exponente cambia un poco, y el cambio producido en el factor exponencial es muchísimo más importante que cualquier otro cambio que pueda ocurrir en los factores que hubiese delante. ¿Por qué puede ser que haya cambios en factores como V_a? Porque nuestro análisis sólo fue aproximado. Después de todo, no hay realmente un volumen definido para cada molécula; a medida que cambiamos la temperatura, el volumen V_a no permanece constante —el líquido se dilata—. Hay otras pequeñas características parecidas y la situación real es más complicada. Hay en todo el volumen factores que dependen de la temperatura y que varían muy despacio. De hecho, podríamos decir que la misma W varía un poquito con la temperatura, ya que a una temperatura más alta y con un volumen molecular diferente, habría distintas atracciones promedio, etc. Así, aunque pensáramos que tener una fórmula en la que todo varía de un modo desconocido con la temperatura, es como no tener fórmula, si nos damos cuenta que el exponente W/kT es en general muy grande, vemos que en la curva de la densidad de vapor en función de la temperatura la mayor parte de la variación es ocasionada por el factor exponencial, y el tomar W como constante y el coeficiente $1/V_a$ casi como constante, es una buena aproximación para pequeños intervalos de la curva. La naturaleza general de la mayor parte de la variación es, en otras palabras, $e^{-W/kT}$.

Resulta que hay muchos, muchos fenómenos en la naturaleza que se caracterizan porque tienen que tomar prestada energía de algún lugar y en los que el distintivo central de la variación de temperatura es e elevado a menos la energía sobre kT. Éste es un hecho muy útil solamente cuando la energía es grande comparada con kT, por lo que la mayor parte de la variación está contenida en la modificacion que se da en kT y no en la constante u otros factores.

Consideremos ahora otro modo de obtener un resultado bastante similar para la evaporación, pero estudiándolo con más detalle. Para llegar a (42.1) simplemente aplicamos una regla que es válida en el equilibrio, pero con el fin de entender las cosas mejor, no hay ningún inconveniente en tratar de considerar los detalles de lo que está ocurriendo. También podemos describir lo que está ocurriendo del modo siguiente: las moléculas del vapor están continuamente bombardeando la superficie del líquido; cuando lo hacen, pueden ser rechazadas o pueden quedar pegadas. Hay un factor desconocido para esto —quizá 50 a 50, quizá 10 a 90, no lo sabemos—. Digamos que siempre se quedan pegadas; lo podemos volver a analizar más tarde suponiendo que no siempre se quedan pegadas. Puesto así, en algún momento habrá un cierto número de átomos que se están condensando en la superficie del líquido. El número de las moléculas que se condensan, el número que llega por unidad de área, por unidad de tiempo, es el número n, las que hay por unidad de volumen, por la velocidad v. Esta velocidad

de las moléculas está relacionada con la temperatura, ya que sabemos que $\frac{1}{2}mv^2$ es igual a $\frac{3}{2}kT$ en promedio. Por lo que v es una especie de velocidad media. Naturalmente deberíamos integrar sobre los ángulos y obtener algo parecido a un promedio, pero es aproximadamente proporcional a la raíz cuadrada de la media de la velocidad al cuadrado, por algún otro factor. Así,

$$N_c = nv \tag{42.2}$$

es la proporción que llega por unidad de área y que se condensa.

Al mismo tiempo, sin embargo, los átomos del líquido están vagando y de vez en cuando uno de ellos es empujado fuera. Ahora tenemos que estimar con qué rapidez son empujados fuera. La idea será que en el equilibrio, el número de los que salen por segundo es igual al número de los que llegan por segundo.

¿Cuántos son empujados fuera? Para que una molécula salga, tiene que haber adquirido accidentalmente un exceso de energía sobre sus vecinos —un exceso de energía considerable, ya que es atraída muy fuertemente por las otras moléculas del líquido—. Ordinariamente no sale debido a esta fuerte atracción, pero algunas veces una de ellas adquiere accidentalmente una energía adicional en las colisiones. Y la probabilidad de que adquiera la energía adicional W que se necesita en nuestro caso es muy pequeña si $kT \ll W$. De hecho, $e^{-W/kT}$ es la probabilidad de que un átomo haya tomado más de esa cantidad de energía. Este es el principio general de la teoría cinética: para que se pueda tomar prestada una energía adicional W sobre el promedio, las probabilidades son e elevado a menos la energía que tenemos que tomar prestada sobre kT. Supongamos ahora que algunas moléculas han tomado prestada esta energía. Tenemos a continuación que estimar cuántas abandonan la superficie por segundo. El que una molécula tenga la energía necesaria no significa que se va a evaporar en realidad, ya que podía estar enterrada muy profundamente en el líquido o, aunque estuviese muy cerca de la superficie, podía estar viajando en una mala dirección. El número que va a abandonar una unidad de área por segundo va a ser algo así: el número de átomos que hay cerca de la superficie, por unidad de área, dividido entre el tiempo que necesita una para escapar, multiplicado por la probabilidad $e^{-W/kT}$ de que estén listos para escapar, entendiendo por esto que tienen energía suficiente.

Supondremos que cada molécula de la superficie del líquido ocupa una sección efectiva de área A. Luego, el número de moléculas por unidad de área de la superficie del líquido será $1/A$. Y ahora, ¿cuánto necesita una molécula para escapar? Si tienen una cierta velocidad promedio v, y tienen que moverse, digamos, un diámetro molecular D, el espesor de la primera capa, el tiempo que les lleve atravesar este espesor es el tiempo necesario para escapar, si la molécula tiene energía suficiente. Será D/v. Así pues, el número de las que se evaporan será aproximadamente

$$N_e = (1/A)(v/D)e^{-W/kT}. \tag{42.3}$$

Ahora bien, el área de cada átomo por el espesor de la capa es aproximadamente lo mismo que el volumen V_a ocupado por un solo átomo. Y por lo tanto, para que pueda haber equilibrio, debemos tener que $N_c = N_e$, es decir,

$$nv = (v/V_a)e^{-W/kT}. \tag{42.4}$$

Podemos cancelar las v puesto que son iguales, aunque una sea la velocidad de una molécula en el vapor y la otra la de una molécula que se está evaporando; son iguales, ya que sabemos que su energía cinética media (en una dirección) es $\frac{1}{2}kT$. Pero alguien puede objetar: "¡No! ¡No! Éstas son especialmente las más rápidas; son las que han tomado energía adicional". Realmente no es así, ya que en el momento en que empiezan a separarse del líquido tienen que *perder* esta energía adicional en favor de la energía potencial. Por ello, a medida que llegan a la superficie, ¡reducen su velocidad a v! Es lo mismo que vimos en nuestra discusión de la distribución de velocidades moleculares en la atmósfera —abajo, las moléculas tienen una cierta distribución de energía—. Las que llegan arriba

tienen la *misma* distribución de energía, puesto que las lentas no llegaron y las rápidas perdieron velocidad. Las moléculas que se están evaporando tienen la misma distribución de energía que las que están dentro —un hecho bastante notable—. De todos modos, es inútil seguir discutiendo con tanto detalle nuestra fórmula a causa de otras inexactitudes, como la probabilidad de ser rechazadas en lugar de entrar en el líquido, etc. Tenemos, pues, una idea vaga de la rapidez de evaporación y condensación y vemos, naturalmente, que la densidad de vapor n varía en la misma forma que antes, pero ahora lo hemos comprendido en más detalle en lugar de tenerlo como una fórmula arbitraria.

Esta comprensión más profunda nos permite analizar algunas cosas. Por ejemplo, supongan que bombeamos el vapor a una velocidad tal que sacamos el vapor en cuanto se forma; si tuviésemos bombas buenas y el líquido se evaporase muy despacio, ¿con qué rapidez ocurriría la evaporación si mantuviésemos el líquido a una temperatura T? Supongan que medimos antes experimentalmente la densidad de vapor en equilibrio; por lo que sabemos, a la temperatura dada, ¿cuántas moléculas por unidad de volumen están en equilibrio con el líquido? Ahora querríamos saber con *qué rapidez* se evaporará. Aunque solamente hemos utilizado un análisis aproximado en lo que concierne a la evaporación, el número de moléculas de vapor que *llegaban* no se estudió tan mal, aparte del factor desconocido del coeficiente de reflexión. Así, podemos usar el hecho de que el número de las que están saliendo, en el equilibrio, es igual al número de las que llegan. Es verdad que estamos quitando el vapor, por lo que sólo hay moléculas que salen, pero si no se tocase el vapor se alcanzaría la densidad de equilibrio para la cual el número de las que vuelven sería igual al número de las que se están evaporando. En resumen, podemos ver fácilmente que el número de las que abandonan la superficie por segundo es igual a uno menos el coeficiente de reflexión desconocido R por el número de las que volverían a la superficie por segundo si el vapor estuviese presente, porque así es como se balancearía la evaporación en el equilibrio:

$$N_e = nv(1-R) = (v(1-R)/V_a)e^{-W/kT}. \tag{42.5}$$

El número de moléculas provenientes del vapor que chocan con la superficie del líquido es, naturalmente, fácil de calcular, ya que no necesitamos conocer tanto acerca de las fuerzas como cuando nos preocupamos de cómo logran escapar a través de la superficie del líquido; es mucho más fácil hacer el razonamiento en el orden inverso.

42-2 Emisión termoiónica

Podemos dar otro ejemplo de una situación muy práctica que es semejante a la evaporación de un líquido —tan semejante que no vale la pena hacer un análisis separado. Es esencialmente el mismo problema—. En una válvula de radio hay una fuente de electrones, un filamento de tungsteno calentado y una placa cargada positivamente para atraer a los electrones. Cualquier electrón que escapa de la superficie de tungsteno es inmediatamente arrastrado a la placa. Ésta es nuestra “bomba” ideal, que está “bombeando” fuera los electrones todo el tiempo. Ahora la pregunta es: ¿cuántos electrones pueden salir por segundo de un pedazo de tungsteno y cómo varía este número con la temperatura? La respuesta a este problema es la misma de (42.5), porque resulta que en un pedazo de metal los electrones son atraídos hacia los iones, o hacia los átomos del metal. Podemos decir *grosso modo* que son atraídos hacia el metal. Para que un electrón salga de un pedazo de metal se necesita una cierta cantidad de energía o trabajo para sacarlo. Este trabajo es diferente para distintas clases de metales. De hecho, varía incluso con el estado de la superficie de una clase dada de metal, pero el trabajo total puede ser de unos pocos elecronvoltios, que es, entre paréntesis, típico de la energía involucrada en reacciones químicas. Podemos considerar este último hecho recordando que el voltaje en una celda química tal como una pila de linterna, que se produce mediante reacciones químicas, es de alrededor de un voltio.

¿Cómo podemos saber cuántos electrones salen por segundo? Sería bastante difícil analizar los efectos en los electrones que salen, es más fácil analizar la situación de modo inverso. Así, podríamos comenzar imaginando que no nos llevásemos los electrones y que ellos fuesen como un gas y pudiesen volver al metal. Luego, habría una cierta densidad de electrones en equilibrio que por supuesto estaría dada por exactamente la misma fórmula que (42.1), donde V_a es el volumen por electrón en el metal, aproximadamente, y W es igual a $q_e\phi$, donde ϕ es el llamado *potencial de arranque*, o sea el voltaje que se necesita para sacar un electrón de la superficie. Esto nos diría cuántos electrones tendría que haber en el espacio circundante incidiendo sobre el metal para balancear a los que están saliendo. Y entonces es fácil calcular cuántos están saliendo si quitamos los que rodean el metal, ya que el número de los que salen es exactamente igual al número de los que incidirían con la densidad anterior del "vapor" de electrones. En otras palabras, la respuesta es que la corriente de electricidad que entra por unidad de área es igual a la carga de cada uno por el número que llega por segundo por unidad de área, lo cual es el número por unidad de volumen por la velocidad, como hemos visto muchas veces:

$$I = q_e nv = (q_e v/V_a)e^{-q_e\phi/kT}. \tag{42.6}$$

Ahora bien, un electronvoltio corresponde a kT, a una temperatura de 11 600 grados. El filamento del tubo puede estar trabajando a una temperatura, digamos, de 1 100 grados, por lo que el factor exponencial es algo así como e^{-10}; cuando variamos la temperatura un poco, el factor exponencial cambia mucho. De nuevo, tenemos que la característica central de la fórmula es $e^{-q_e\phi/kT}$. En realidad, el factor que va delante está totalmente errado —resulta que el comportamiento de los electrones en un metal no se describe correctamente mediante la teoría clásica, sino mediante la mecánica cuántica, pero esto no hace sino cambiar un poco dicho factor—. En realidad, nadie ha sido capaz de obtener correctamente esto, aunque muchos han usado la mecánica cuántica de alta categoría para sus cálculos. El gran problema es, ¿cambia W ligeramente con la temperatura? Si es así, no se puede distinguir una W que cambia lentamente con la temperatura, de un coeficiente diferente delante. Esto es, si W cambia linealmente, digamos, con la temperatura, de manera que $W = W_0 + \alpha kT$, tendríamos

$$e^{-W/kT} = e^{-(W_0+\alpha kT)/kT} = e^{-\alpha}e^{-W_0/kT}.$$

O sea, una W que depende linealmente de la temperatura es equivalente a una "constante" diferente. Es realmente muy difícil y poco fructífero tratar de obtener exactamente el coeficiente que va delante.

42-3 Ionización térmica

Veamos ahora otro ejemplo de la misma idea, siempre la misma idea. Tiene que ver con la ionización. Supongan que tenemos en un gas un montón de átomos en estado neutro, digamos, pero que el gas está caliente y los átomos se pueden ionizar. Querríamos saber cuántos iones hay en una circunstancia determinada si tenemos una cierta densidad de átomos por unidad de volumen a una cierta temperatura. Consideramos de nuevo una caja en la que hay N átomos que pueden contener electrones. (Si un electrón ha salido de un átomo, éste se llama *ion* y si el átomo es neutro se llamará simplemente átomo.) Supongan entonces que, en un momento dado, el número de átomos neutros es n_a, el número de iones es n_i y el número de electrones es n_e, todos por unidad de volumen. El problema es: ¿cuál es la relación entre estos tres números?

En primer lugar, tenemos dos condiciones o restricciones para los números. Por ejemplo, según variemos diferentes condiciones, como la temperatura u otras cosas, $n_a + n_i$ permanecerá constante, ya que esto es simplemente el número N de núcleos atómicos que hay en la caja. Si conservamos fijo el número de núcleos por unidad de volumen y cambiamos, por ejemplo, la temperatura, a medida que la ionización tiene lugar algunos átomos pasarán a iones, pero el número

total de átomos más iones será constante. Esto es, $n_a + n_i = N$. Otra condición es que si el gas total tiene que estar eléctricamente neutro (y no tenemos en cuenta una doble o triple ionización), esto significa que el número de iones es igual al número de electrones para cualquier tiempo, o sea $n_i = n_e$. Éstas son ecuaciones subsidiarias que expresan simplemente la conservación de la carga y la conservación de los átomos.

Estas ecuaciones son ciertas y las usaremos finalmente cuando consideremos un problema real. Pero queremos obtener otra relación entre las cantidades. Lo podemos hacer como sigue. Usamos de nuevo la idea de que se necesita una cierta cantidad de energía para sacar al electrón del átomo, que llamaremos *energía de ionización*, y la escribiremos como W para que todas las fórmulas se parezcan. Así, W es la energía necesaria para sacar un electrón de un átomo y convertirlo en un ion. Ahora decimos de nuevo que el número de electrones libres por unidad de volumen en el "vapor" es igual al número de electrones ligados a los átomos por unidad de volumen por e elevado a menos la diferencia de energía entre estar ligado y estar libre, sobre kT. Otra vez tenemos la ecuación básica. ¿Cómo podemos escribirla? El número de electrones libres por unidad de volumen sería, naturalmente, n_e, ya que ésta es la definición de n_e. Ahora bien, ¿qué hay del número de electrones ligados a los átomos por unidad de volumen? El número total de lugares donde podríamos colocar los electrones es aparentemente $n_a + n_i$ y supondremos que cuando están ligados, cada uno de ellos lo está dentro de un cierto volumen V_a. Luego, la cantidad de volumen disponible para los electrones que estarían ligados es $(n_a + n_i)V_a$ y de ese modo podríamos querer escribir nuestra fórmula en la forma

$$n_e = \frac{n_a}{(n_a + n_i)V_a}\, e^{-W/kT}.$$

Sin embargo, la fórmula es errónea en un punto esencial, que es el siguiente: cuando ya hay un electrón en un átomo, ¡otro electrón no puede venir jamás a ese volumen! En otras palabras, todos los volúmenes de todos los lugares posibles no están realmente disponibles para aquel electrón que está tratando de decidir si estar o no en el vapor o en la posición condensada, porque en este problema hay una característica adicional: cuando un electrón se encuentra donde está otro electrón no se le permite ir —es repelido—. Por esta razón, resulta que deberíamos contar solamente aquella parte del volumen que está disponible para que un electrón se quede o no. Esto es, las partes que ya están ocupadas no cuentan en el volumen total disponible, sino que el único volumen que se permite es el de los *iones*, donde hay lugares vacíos para que el electrón vaya. Luego, en estas circunstancias, encontramos que un modo mejor de escribir nuestra fórmula es

$$\frac{n_e n_i}{n_a} = \frac{1}{V_a}\, e^{-W/kT}. \tag{42.7}$$

Esta fórmula se llama *ecuación de ionización de Saha*. Veamos ahora si podemos entender cualitativamente por qué una fórmula como ésta es cierta, haciendo uso de las cosas cinéticas que están ocurriendo.

En primer lugar, de vez en cuando un electrón llega a un ion y se combinan para formar un átomo. Y también, de vez en cuando, un átomo sufre una colisión y se separa en un ion y un electrón. La rapidez de ambos procesos debe ser igual. ¿Con qué rapidez se encuentran los electrones y los iones? La rapidez aumenta ciertamente si se incrementa el número de electrones por unidad de volumen. También se incrementa si aumenta el número de iones por unidad de volumen. Esto es, la rapidez total a la cual ocurre la recombinación es ciertamente proporcional al número de electrones por el número de iones. Ahora bien, la rapidez total a la que está ocurriendo la ionización debido a las colisiones debe depender linealmente del número de átomos que hay para ser ionizados. Por consiguiente, ambas rapideces se balancearán cuando haya alguna relación entre el producto $n_e n_i$ y el número de átomos n_a. El hecho de que suceda que esta relación está dada por esta fórmula particular, donde W es la energía de ionización, es

naturalmente algo más de información, pero podemos entender fácilmente que la fórmula contendría necesariamente las concentraciones de los electrones, iones y átomos en la combinación $n_e n_i/n_a$ para producir una constante independiente de las n, y dependiente solamente de la temperatura, de las secciones eficaces atómicas y otros factores constantes.

También podemos notar que, puesto que en la ecuación intervienen los números *por unidad de volumen*, si tuviésemos que hacer dos experimentos con un número total N de átomos más iones, o sea con cierto número fijo de núcleos, pero usando cajas de volúmenes diferentes, las n serían más pequeñas en la caja más grande. Pero, puesto que el cociente $n_e n_i/n_a$ permanece igual, el *número total* de iones y electrones debe ser mayor en la caja más grande. Para ver esto, supongan que hay N núcleos dentro de una caja de volumen V y que una fracción f de ellos está ionizada. Entonces, $n_e = fN/V = n_i$ y $n_a = (1-f)N/V$. Nuestra ecuación se convierte en

$$\frac{f^2}{1-f}\,\frac{N}{V} = \frac{e^{-W/kT}}{V_a}. \tag{42.8}$$

En otras palabras, si tomamos una densidad de átomos cada vez más pequeña, o hacemos el volumen donde están encerrados cada vez mayor, la fracción f de iones y electrones debe aumentar. Esta ionización, sólo por "expansión" a medida que la densidad disminuye, es la razón por la cual creemos que a densidades muy bajas, como en el espacio frío entre las estrellas, puede haber iones presentes, aunque no lo podamos entender desde el punto de vista de la energía disponible. Aunque se necesitan muchos, muchos kT de energía para hacerlos, hay iones presentes.

¿Por qué puede haber iones presentes cuando hay tanto espacio alrededor, mientras que si aumentamos la densidad los iones tienden a desaparecer? *Respuesta*: Consideren un átomo. De vez en cuando, luz, u otro átomo, o un ion o cualquier cosa que sea la que mantiene el equilibrio térmico, lo golpea. Muy raramente, ya que se necesita una cantidad tremenda de energía adicional, un electrón sale y queda un ion. Ahora bien, si el espacio es enorme, ese electrón vaga y vaga y quizás no llega cerca de algo en años. Pero alguna vez en un tiempo muy grande, sí vuelve a un ion y se combinan para formar un átomo. Por lo que la rapidez a la que los electrones salen de los átomos es muy pequeña. Pero si el volumen es enorme, un electrón que ha escapado toma tanto tiempo en encontrar otro ion para recombinarse, que su probabilidad de recombinación es muy, pero muy pequeña; así, a pesar de la gran energía adicional que se necesita, puede haber un número razonable de electrones.

42-4 Cinética química

La misma situación que acabamos de llamar "ionización" se encuentra también en una reacción química. Por ejemplo, si dos objetos A y B se combinan para formar un compuesto AB; entonces, si pensamos por un momento, vemos que AB es lo que hemos llamado un átomo, B lo que llamamos un electrón y A lo que llamamos un ion. Con estas sustituciones, las ecuaciones de equilibrio tienen exactamente la misma forma:

$$\frac{n_A n_B}{n_{AB}} = ce^{-W/kT}. \tag{42.9}$$

Desde luego, esta fórmula no es exacta puesto que la "constante" c depende del volumen que se permita a A y B para combinarse, entre otras cosas, pero mediante razonamientos termodinámicos se puede identificar el significado de W en el factor exponencial, y resulta que es muy próximo a la energía que se necesita en la reacción.

Supongan que hemos intentado entender esta fórmula como resultado de colisiones, muy parecido al modo como entendimos la fórmula de evaporación, argumentando acerca de cuántos electrones salían y cuántos volvían por unidad de tiempo. Supongan que A y B se combinan de vez en cuando mediante una

colisión para formar un compuesto AB. Y supongan que el compuesto AB es una molécula complicada que camina al azar y que es golpeada por otras moléculas, y que de vez en cuando obtiene energía suficiente para explotar y romperse de nuevo en A y B.

Ahora bien, en las reacciones químicas resulta realmente que si los átomos se acercan con energías muy pequeñas, aun cuando se pueda emplear alguna energía en la relación $A + B \to AB$, el hecho de que A y B se puedan tocar no hace que la reacción comience necesariamente. Se requiere usualmente que la colisión sea bastante fuerte, de hecho, para hacer que la reacción comience —una colisión "suave" entre A y B puede no originarla, aunque se utilice alguna energía en el proceso—. Así pues, supongamos que es muy común en reacciones químicas que, para que de A y B se forme AB, no basta que choquen entre sí, sino que tienen que hacerlo *con energía suficiente*. Esta energía se llama *energía de activación* —energía que se necesita para "activar" la reacción—. Llamemos A^* a la energía de activación, energía adicional necesaria en una colisión para que la reacción pueda ocurrir realmente. Entonces, en la velocidad R_f a la que A y B produce AB intervendría el número de átomos de A por el número de átomos de B, por la velocidad a la cual un solo átomo golpearía alguna sección eficaz σ_{AB}, por un factor $e^{-A^*/kT}$, que es la probabilidad de que tengan energía suficiente:

$$R_f = n_A n_B v \sigma_{AB} e^{-A^*/kT}. \tag{42.10}$$

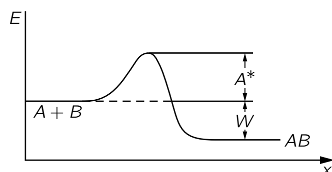


Figura 42-1. Relación de energía para la reacción $A + B \to AB$.

Ahora tenemos que encontrar la velocidad opuesta R_r. Hay una cierta probabilidad de que AB se desintegre. Para que pueda hacerlo, no solamente debe tener la energía W que necesita para que se separen, sino que así como era difícil que A y B se unieran, hay una especie de montaña sobre la cual A y B tienen que pasar para separarse de nuevo; no sólo deben tener bastante energía para estar en condiciones de separarse, sino además una cierta energía adicional. Es como trepar una montaña para llegar a un valle profundo; tienen que trepar la montaña para entrar y tienen que trepar para salir del valle y volver atrás (figura 42-1). Por lo tanto, la velocidad con la que AB vuelve a A y B será proporcional al número n_{AB} que hay presente, multiplicado por $e^{-(W+A^*)/kT}$.

$$R_r = c' n_{AB} e^{-(W+A^*)/kT}. \tag{42.11}$$

La c' incluirá el volumen de los átomos y las colisiones por unidad de tiempo (que podríamos calcular con áreas, tiempos y espesores, como hicimos en el caso de evaporación; pero no lo haremos). La característica principal de interés para nosotros es que cuando estas dos velocidades son iguales, su cociente nos da la unidad. Esto nos dice que $n_A n_B / n_{AB} = c e^{-W/kT}$, igual que antes, donde c involucra las secciones efectivas, las velocidades y otros factores independientes de las n.

Lo interesante es que la velocidad de la reacción también varía como una exponencial, $e^{-\text{constante}/kT}$; aunque la constante no es la misma que la que rige las concentraciones, la energía de activación A^* es muy diferente de la energía W. *W rige las proporciones de A, B y AB que tenemos en equilibrio*, pero si queremos saber con qué velocidad $A + B$ pasa a AB, esto no es una cuestión de equilibrio, y por consiguiente una energía diferente, *la energía de activación*, rige la velocidad de reacción mediante un factor exponencial.

Además, A^* no es una constante fundamental como W. Supongan que en la superficie de la pared —o en algún otro lugar— A y B se pudiesen adherir temporalmente de tal modo que pudiesen combinarse más fácilmente. En otras palabras, podríamos encontrar un "túnel" a través de la montaña o quizá una montaña más baja. Por conservación de la energía, cuando esté todo terminado, habremos hecho AB a partir de A y B, por lo que la diferencia de energía W será enteramente independiente del modo en que ocurrió la reacción, pero la energía de activación, A^*, dependerá *muchísimo* del modo en que ocurre la reacción. Esto explica por qué las reacciones químicas son muy sensibles a las condiciones exteriores. Podemos cambiar la velocidad colocando una superficie de una clase

diferente, podemos colocarla en un "barril diferente", por ejemplo, y ocurrirá a una velocidad diferente, si ésta depende de la naturaleza de la superficie. O si colocamos dentro un tercer objeto diferente, puede cambiar muchísimo la velocidad; algunas cosas producen enormes cambios en la velocidad simplemente cambiando A^* un poquito —se les llama *catalizadores*—. Puede ocurrir que una reacción prácticamente no se lleve a cabo porque A^* es demasiado grande a la temperatura dada, pero cuando introducimos esta sustancia especial, el catalizador, entonces la reacción es verdaderamente rápida porque A^* se ha reducido.

Entre paréntesis, hay alguna dificultad con tal reacción, A más B da AB, porque no podemos conservar a la vez energía y momentum cuando tratamos de colocar juntos dos objetos para formar uno que es más estable. Por consiguiente, necesitamos por lo menos un tercer objeto C y por ello la verdadera reacción es mucho más complicada. En la velocidad hacia adelante interviene el producto $n_A n_B n_C$ y podría parecer que nuestra fórmula va a ser errónea, pero ¡no! Cuando consideramos la velocidad a la que AB va en la otra dirección, encontramos que también necesita colisionar con C, por lo que en la velocidad inversa interviene $n_{AB} n_C$, las n_C se cancelan en la fórmula para las concentraciones de equilibrio. La ley de equilibrio (42.9) que escribimos primero está absolutamente garantizada como verdadera ¡cualquiera que sea el mecanismo de la reacción!

42-5 Leyes de radiación de Einstein

Volvemos ahora a una interesante situación análoga que tiene que ver con la ley de radiación de cuerpo negro. En el capítulo anterior calculamos la ley de distribución para la radiación en una cavidad del modo que Planck lo hizo, considerando la radiación de un oscilador. El oscilador tiene que contar con una cierta energía media, y como está oscilando, radiaría y seguiría bombeando radiación a la cavidad hasta que amontonase tanta que balancease la absorción y emisión. Así encontramos que la intensidad de radiación a la frecuencia ω estaba dada por la fórmula

$$I(\omega)\,d\omega = \frac{\hbar\omega^3\,d\omega}{\pi^2 c^2 (e^{\hbar\omega/kT} - 1)}. \tag{42.12}$$

Este resultado contenía la suposición de que el oscilador que generaba la radiación tenía niveles de energía definidos e igualmente espaciados. No dijimos que la luz tenía que ser un fotón o algo parecido. No discutimos acerca de cómo, cuando un átomo pasa de un nivel a otro, la energía tiene que salir en una unidad de energía, $\hbar\omega$, en forma de luz. La idea original de Planck fue que era la materia la que estaba cuantizada y no la luz; los osciladores materiales no pueden tomar simplemente cualquier energía, sino que lo tienen que hacer en bloques. Además, el problema con la derivación es que era parcialmente clásica. Calculamos la rapidez de radiación de un oscilador de acuerdo con la física clásica, después le dimos la vuelta y dijimos: "no, este oscilador tiene muchos niveles de energía". Así que gradualmente, con el fin de encontrar el resultado correcto, el resultado completamente cuántico, hubo un desarrollo lento que culminó en la mecánica cuántica de 1927. Pero, entre tanto, Einstein intentó convertir el punto de vista de Planck de que solamente los osciladores de materia estaban cuantizadas, a la idea de que la luz era realmente fotones y podía ser considerada de algún modo como partículas con energía $\hbar\omega$. Además, Bohr había indicado que *cualquier* sistema de átomos tiene niveles de energía, pero que éstos no se encuentran necesariamente igualmente espaciados como el oscilador de Planck. Así que fue necesario volver a derivar o por lo menos volver a discutir la ley de radiación desde un punto de vista cuántico más completo.

Einstein supuso que la fórmula de Planck estaba bien y la usó para obtener alguna información nueva, desconocida anteriormente, acerca de la interacción de la radiación con la materia. Su discusión fue como sigue: consideren cualesquiera dos de los muchos niveles de energía de un átomo, digamos los niveles m-ésimo y n-ésimo (figura 42-2). Ahora bien, Einstein propuso que cuando luz de frecuencia

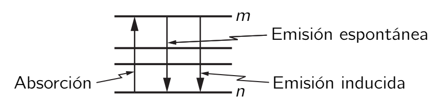


Figura 42-2. Transiciones entre dos niveles de energía de un átomo.

apropiada incide sobre ese átomo, éste puede absorber ese fotón de luz y efectuar una transición del estado n al m, y que la probabilidad de que ocurra así por segundo depende, naturalmente, de los dos niveles, pero es *proporcional a la intensidad de la luz* que está incidiendo. Llamemos a esta constante de proporcionalidad B_{mn}, simplemente para tener presente que no es una constante universal de la naturaleza, sino que depende del par particular de niveles, algunos niveles son fáciles de excitar y otros difíciles. Ahora bien, ¿cuál va a ser la fórmula para la probabilidad de emisión de m a n? Einstein propuso que debe constar de dos partes. En primer lugar, aun cuando no hubiese luz presente, habría alguna probabilidad de que un átomo en un estado excitado pasase a un estado más bajo emitiendo un fotón, a esto lo llamaremos *emisión espontánea*. Es análogo a la idea de que un oscilador con una cierta cantidad de energía, aun en física clásica, no la guarda sino que la pierde por radiación. Así, lo análogo de la radiación espontánea de un sistema clásico es que si un átomo está en un estado excitado hay una cierta probabilidad A_{mn}, que también depende de los niveles, de que baje de m a n y esta probabilidad es independiente de si hay luz iluminando el átomo o no. Pero entonces Einstein fue más lejos y mediante la comparación con la teoría clásica y otros argumentos concluyó que la emisión también se veía influenciada por la presencia de luz —en el momento que la luz de frecuencia apropiada ilumina el átomo, éste tiene una probabilidad mayor de emitir un fotón, la cual es proporcional a la intensidad de la luz, con la constante de proporcionalidad B_{mn}—. Más tarde, si deducimos que este coeficiente es cero, habremos demostrado que Einstein estaba equivocado. Pero, naturalmente, encontraremos que tenía razón.

Así, Einstein supuso que hay tres clases de procesos: una absorción proporcional a la intensidad de la luz, una emisión proporcional a la intensidad de la luz llamada *emisión inducida* y algunas veces *emisión estimulada*, y una emisión espontánea independiente de la luz.

Supongan ahora que tenemos, en el equilibrio a una temperatura T, un cierto número de átomos N_n en el estado n y otro número N_m en el estado m. Entonces, el número total de átomos que van a pasar de n a m es el número de los que están en el estado n por la probabilidad por segundo de que si uno está en n pase a m. La fórmula para el número de los que van a pasar de n a m por segundo es:

$$R_{n\to m} = N_n B_{nm} I(\omega). \tag{42.13}$$

Del mismo modo, el número de los que pasarán de m a n se expresa, como el número N_m de los que están en m, multiplicado por la probabilidad por segundo de que cada uno baje a n. La expresión es

$$R_{m\to n} = N_m[A_{mn} + B_{mn} I(\omega)]. \tag{42.14}$$

Supondremos ahora que en el equilibrio térmico el número de átomos que suben debe ser igual al número de los que bajan. Por lo menos es un modo de asegurarse de que el número de átomos permanezca constante en cada nivel.* Consideramos entonces que estas dos probabilidades son iguales en el equilibrio. Pero tenemos aún otra información: sabemos lo grande que es N_m comparado con N_n —su cociente es $e^{-(E_m-E_n)/kT}$—. Ahora, Einstein supuso que la única luz que efectivamente produce la transición de n a m es aquella que tiene la frecuencia correspondiente a la diferencia de energía, por lo que $E_m - E_n = \hbar\omega$ en todas nuestras fórmulas. Entonces

$$N_m = N_n e^{-\hbar\omega/kT}. \tag{42.15}$$

Y si igualamos las dos probabilidades: $N_n B_{nm} I(\omega) = N_m[A_{mn} + B_{mn} I(\omega)]$ y dividimos entre N_m, obtenemos

* Éste no es el único modo de arreglárnoslas para mantener constante el número de átomos en los diversos niveles, pero es el que funciona. El hecho de que en el equilibrio térmico cada proceso deba ser balanceado por su opuesto exacto es lo que se llama *principio de balance detallado*.

$$B_{nm}I(\omega)e^{\hbar\omega/kT} = A_{mn} + B_{mn}I(\omega). \tag{42.16}$$

A partir de esta ecuación podemos calcular $I(\omega)$. Es simplemente

$$I(\omega) = \frac{A_{mn}}{B_{nm}e^{\hbar\omega/kT} - B_{mn}}. \tag{42.17}$$

Pero Planck nos ha dicho antes que la fórmula debe ser (42.12). Por tanto podemos deducir algo: en primer lugar, B_{nm} debe ser igual a B_{mn}, puesto que de otra forma no podemos obtener $(e^{\hbar\omega/kT} - 1)$. Así, Einstein descubrió algunas cosas que no supo cómo calcular, en especial *que la probabilidad de emisión inducida y la probabilidad de absorción deben ser iguales.* Esto es interesante. Y además, para que (42.17) y (42.12) concuerden,

$$A_{mn}/B_{mn} \qquad \text{debe ser} \qquad \hbar\omega^3/\pi^2c^2. \tag{42.18}$$

Así, por ejemplo, si conocemos la probabilidad de absorción para un nivel dado, podemos deducir la probabilidad de emisión espontánea y la de emisión inducida o cualquier combinación.

Hasta aquí es hasta donde Einstein o cualquier otro podía llegar usando tales razonamientos. Hoy en día, computar la probabilidad absoluta de emisión espontánea o las otras probabilidades para cualquier transición atómica específica requiere, naturalmente, un conocimiento de la maquinaria del átomo, llamada electrodinámica cuántica, que no se descubrió hasta 11 años más tarde. Este trabajo de Einstein se hizo en 1916.

La posibilidad de emisión inducida ha encontrado actualmente aplicaciones interesantes. Si hay luz presente tenderá a inducir la transición hacia abajo. La transición añade su $\hbar\omega$ a la energía de luz disponible, si hay algunos átomos situados en el estado superior. Podemos hacer, mediante algún método no termodinámico, que un gas tenga el número en el estado m mucho más grande que el número en el estado n. Esto está muy lejos del equilibrio y no está dado entonces por la fórmula $e^{-\hbar\omega/kT}$ del equilibrio. También podemos hacer que el número en el estado superior sea muy grande, mientras que el del estado inferior sea prácticamente cero. En este caso, la luz que tiene la energía correspondiente a la diferencia de energías $E_m - E_n$ no será fuertemente absorbida, ya que no hay muchos átomos en el estado n que la puedan absorber. Por otro lado, cuando esta luz está presente, ¡inducirá la emisión desde este estado superior! Por lo tanto, si tenemos muchos átomos en el estado superior, habrá una especie de reacción en cadena en la cual, desde el momento en que los átomos comiencen a emitir, otros más se verán forzados a emitir y todo el conjunto de átomos caerá de golpe. Esto es lo que se llama un *láser* o, en el caso del infrarrojo lejano, un *máser de amoniaco.*

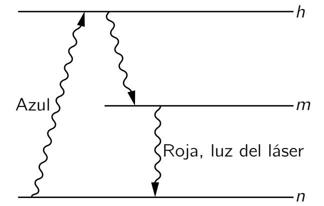


Figura 42-3. Excitando un estado más alto h, mediante luz azul digamos, que puede emitir un fotón dejando átomos en el estado m, el número en este estado m se hace lo suficientemente grande como para comenzar la acción láser.

Se pueden usar diversos artificios para obtener los átomos en el estado m. Podría haber niveles más altos a los cuales los átomos podrían ir si iluminásemos con un haz fuerte de luz de alta frecuencia. De estos niveles podrían gotear emitiendo varios fotones, hasta que todos se queden parados en el estado m. Si tienden a permanecer en el estado m sin emitir, el estado se llama *átomo metaestable.* Y luego se hace que todos caigan de golpe por emisiones inducidas. Otro punto técnico más, si colocamos este sistema en una caja ordinaria, radiaría espontáneamente en tantas direcciones diferentes, comparado con el efecto inducido, que aún tendríamos problemas. Pero podemos acrecentar el efecto inducido, aumentar su eficiencia, colocando espejos casi perfectos en cada lado de la caja, de modo que la luz que es emitida tenga otra oportunidad, y otra oportunidad, y otra oportunidad, de inducir más emisión. Aunque los espejos son casi 100 % reflectores, hay una pequeña cantidad de transmisión y un poco de luz sale fuera. Al final, a causa de la conservación de la energía, toda la luz sale uniformemente en una linda línea recta, lo cual da lugar a los haces de luz intensos que son posibles hoy en día con los láseres.

43

Difusión

43-1 Colisiones entre moléculas

Hasta ahora sólo hemos considerado los movimientos moleculares en un gas que se encuentra en equilibrio térmico. Ahora queremos estudiar lo que ocurre cuando se está cerca del equilibrio, pero no exactamente en el equilibrio. En un estado muy apartado del equilibrio, las cosas son extremadamente complicadas, pero muy cerca del equilibrio podemos examinar fácilmente lo que sucede. Para verlo debemos, sin embargo, volver a la teoría cinética. La mecánica estadística y la termodinámica se ocupan del estado de equilibrio, pero fuera del equilibrio sólo podemos analizar lo que ocurre átomo por átomo, por así decir.

Como ejemplo sencillo de una situación de falta de equilibrio, consideraremos la difusión de iones en un gas. Supongan que en un gas hay una concentración relativamente baja de iones —moléculas cargadas eléctricamente—. Si aplicamos un campo eléctrico al gas, sobre cada ion actuará una fuerza que es diferente de las fuerzas entre las moléculas neutras del gas. Si no hubiera otras moléculas presentes, un ion tendría una aceleración constante hasta alcanzar la pared del recipiente. Sin embargo, no lo puede hacer debido a la presencia de las otras moléculas; su velocidad aumenta solamente hasta que choca con una molécula y pierde su momentum. Nuevamente empieza a adquirir mayor velocidad y de nuevo pierde su momentum. El efecto resultante es que un ion se abre camino a lo largo de una trayectoria irregular, aunque con un movimiento neto en la dirección del campo eléctrico. Veremos que el ion experimenta un “arrastre” medio, con una velocidad media proporcional al campo eléctrico —cuanto más intenso es el campo eléctrico, más rápido va—. Mientras el campo está aplicado y mientras el ion se está moviendo, *no* está, por supuesto, en equilibrio térmico; está tratando de llegar al equilibrio, que consiste en quedarse en el extremo del recipiente. Además, podemos calcular la velocidad de arrastre por medio de la teoría cinética.

Resulta que con la matemática que sabemos actualmente, en realidad no podemos calcular *con precisión* lo que sucederá, pero podemos obtener resultados aproximados que muestran los rasgos esenciales. Podemos descubrir cómo varían las cosas con la presión, con la temperatura, etc., pero no será posible obtener exactamente los factores numéricos correctos que hay delante de todos los términos. En consecuencia, en nuestras derivaciones no nos preocuparemos del valor preciso de los factores numéricos. Sólo se los puede obtener con un tratamiento matemático mucho más refinado.

Antes de considerar lo que sucede cuando no hay equilibrio, tendremos que examinar un poco más de cerca lo que pasa en un gas en equilibrio térmico. Tendremos que conocer, por ejemplo, cuál es el tiempo medio entre colisiones sucesivas de una molécula.

Cualquier molécula experimenta una serie de colisiones con otras moléculas —naturalmente que al azar—. En un largo periodo de tiempo T, una molécula determinada sufrirá un cierto número N de choques. Si duplicamos la duración del intervalo habrá el doble de choques. Luego, el número de colisiones es proporcional al tiempo T. Lo escribiremos así:

$$N = T/\tau. \tag{43.1}$$

Hemos escrito la constante de proporcionalidad en la forma $1/\tau$, donde τ tiene dimensiones de tiempo. La constante τ es el tiempo medio entre colisiones.

Supongan, por ejemplo, que hay 60 colisiones en una hora, τ es entonces un minuto. Diríamos que τ (un minuto) es el *tiempo medio* entre las colisiones.

Muchas veces queremos hacernos la siguiente pregunta: "¿cuál es la *probabilidad* de que una molécula experimente una colisión en el siguiente *intervalo pequeño* de tiempo dt?" La respuesta, que la podemos entender intuitivamente, es dt/τ. Pero tratemos de hacer un razonamiento más convincente. Supongan que había un gran número N de moléculas. ¿Cuántas tendrán colisiones en el siguiente intervalo dt? Si hay equilibrio, nada varía *en promedio* con el tiempo. Por lo tanto, N moléculas esperando durante dt sufrirán el mismo número de colisiones que *una* molécula durante el tiempo $N\,dt$. Sabemos que ese número es $N\,dt/\tau$. Luego, el número de choques de N moléculas es $N\,dt/\tau$ en el tiempo dt y la probabilidad de un choque para cualquier molécula es simplemente $1/N$ por ese valor, o sea $(1/N)(N\,dt/\tau) = dt/\tau$, como habíamos supuesto más arriba. Es decir, la fracción de moléculas que sufrirán una colisión en el tiempo dt es dt/τ. Para dar un ejemplo, si τ es un minuto, la fracción de moléculas que sufrirá colisiones en un segundo es $1/60$. Esto significa, naturalmente, que $1/60$ de las moléculas están por casualidad lo bastante cerca de donde van a chocar como para que *sus* colisiones ocurran en el segundo siguiente.

Cuando decimos que τ, el tiempo medio entre colisiones, es un minuto, no queremos decir que todas las colisiones ocurrirán a intervalos exactos de un minuto. Una partícula determinada no tiene una colisión, espera un minuto y tiene otra colisión. Los tiempos entre colisiones sucesivas son completamente variables. No lo necesitaremos para el trabajo que sigue, pero podemos hacer una pequeña digresión para contestar la pregunta: "¿cuáles *son* los tiempos entre colisiones?" Sabemos que para el caso anterior, el tiempo *medio* es un minuto, pero nos gustaría saber, por ejemplo, cuál es la probabilidad de que no tengamos ninguna colisión durante *dos* minutos.

Encontraremos la respuesta a la pregunta general: "¿cuál es la probabilidad de que una molécula ande durante un tiempo t sin tener una colisión?" En un instante arbitrario —que llamamos $t = 0$— comenzamos a observar una molécula determinada. ¿Cuál es la probabilidad de que llegue hasta t sin chocar con otra molécula? Para calcular esta probabilidad, observamos lo que está ocurriendo con todas las N_0 moléculas dentro de un recipiente. Después de haber esperado un tiempo t, encontraremos que algunos habrán tenido colisiones. Sea $N(t)$ el número que *no* ha tenido colisiones hasta el instante t. $N(t)$ es, por supuesto, menor que N_0. Podemos hallar $N(t)$ porque sabemos cómo varía en el tiempo. Si sabemos que $N(t)$ moléculas han llegado hasta t, el número que llega hasta $t+dt$, $N(t+dt)$, es *menor* que $N(t)$ en el número que ha tenido colisiones durante dt. Hemos escrito más arriba el número de colisiones en dt en función del tiempo medio τ como $dN = N(t)dt/\tau$. Tenemos la ecuación:

$$N(t+dt) = N(t) - N(t)\,\frac{dt}{\tau}. \tag{43.2}$$

La cantidad del primer miembro, $N(t+dt)$, se puede escribir, de acuerdo con las definiciones del cálculo infinitesimal, en la forma $N(t) + (dN/dt)dt$. Haciendo esta sustitución, la ecuación (43.2) da:

$$\frac{dN(t)}{dt} = -\frac{N(t)}{\tau}. \tag{43.3}$$

El número de las que se pierden en el intervalo dt es proporcional al número de las presentes e inversamente proporcional a la vida media τ. La ecuación (43.3) se puede integrar fácilmente si la escribimos en la forma:

$$\frac{dN(t)}{N(t)} = -\frac{dt}{\tau}. \tag{43.4}$$

Cada miembro es un diferencial exacto, por lo que la integral es:

$$\ln N(t) = -t/\tau + (\text{una constante}), \tag{43.5}$$

que es lo mismo que

$$N(t) = (\text{constante})e^{-t/\tau}. \tag{43.6}$$

Sabemos que la constante debe ser precisamente N_0, número total de moléculas presentes, puesto que $t = 0$, todas empiezan a esperar su colisión "siguiente". Podemos escribir nuestro resultado en la forma

$$N(t) = N_0 e^{-t/\tau}. \tag{43.7}$$

Si deseamos la *probabilidad* $P(t)$ de que no haya colisiones, la podemos obtener dividiendo $N(t)$ entre N_0:

$$P(t) = e^{-t/\tau}. \tag{43.8}$$

Nuestro resultado es: la probabilidad de que una molécula determinada sobreviva un tiempo t sin sufrir colisiones es $e^{-t/\tau}$, donde τ es el tiempo medio entre colisiones. La probabilidad comienza en 1 (o sea certeza) en $t = 0$ y disminuye a medida que t aumenta. La probabilidad de que la molécula evite colisiones durante un tiempo igual a τ es $e^{-1} \approx 0.37$. Además, la probabilidad de que lo haga durante un tiempo mayor que el tiempo medio es menor que un medio. Eso está muy bien, porque hay bastantes moléculas que antes de chocar no sufren colisiones durante periodos *más largos* que el tiempo medio, por lo que el tiempo medio puede seguir siendo τ.

Al principio definimos τ como el tiempo medio *entre* colisiones. El resultado obtenido en la ecuación (43.7) dice también que el tiempo libre entre un instante inicial *arbitrario* y la colisión *siguiente* es *también* τ. Podemos demostrar este hecho un tanto sorprendente de la manera que sigue. El número de moléculas que sufren su colisión *siguiente* en el intervalo dt a un tiempo t a partir de un instante inicial elegido arbitrariamente es $N(t)dt/\tau$. Su "intervalo de tiempo hasta la siguiente colisión" es precisamente t. El "tiempo medio hasta la colisión siguiente" se obtiene de la manera habitual:

$$\text{Tiempo medio hasta la colisión siguiente} = \frac{1}{N_0}\int_0^\infty t\,\frac{N(t)\,dt}{\tau}.$$

Empleando $N(t)$ obtenido en (43.7) y calculando la integral, encontramos verdaderamente que τ es el tiempo medio desde *cualquier* instante hasta la colisión siguiente.

43-2 El camino libre medio

Otro modo de describir las colisiones moleculares es hablar, no del *tiempo* entre colisiones, sino del *camino* que recorre la partícula entre colisiones. Si decimos que el tiempo medio entre colisiones es τ y que las moléculas tienen una velocidad media v, podemos esperar que la *distancia* media entre colisiones, que llamaremos l, sea simplemente el producto de τ y v. Esta distancia entre colisiones se denomina comúnmente *camino libre medio*:

$$\text{Camino libre medio } l = \tau v. \tag{43.9}$$

En este capítulo seremos un poco imprecisos acerca de *qué clase de promedio* entendemos en cualquier caso particular. Los diversos promedios posibles —el valor medio, el valor medio cuadrático, etc.— son todos casi iguales y difieren en factores cercanos a uno. Como de todas maneras se necesita un análisis detallado para obtener los factores numéricos correctos, no es preciso que nos preocupemos de qué promedio se necesita para cualquier problema en particular. También podemos advertir al lector que los símbolos algebraicos que estamos usando para algunas cantidades físicas (por ejemplo, l para el camino libre medio) no siguen una convención de aceptación general, principalmente porque no hay acuerdo general.

Así como la probabilidad de que una molécula tenga una colisión en un intervalo breve dt es igual a dt/τ; la probabilidad de que tenga una colisión al recorrer una distancia dx es dx/l. Usando los mismos razonamientos que anteriormente, el lector puede demostrar que la probabilidad de que una molécula recorra por lo menos una distancia x antes de tener el choque siguiente es $e^{-x/l}$.

La distancia media que una molécula recorre antes de chocar con otra —el camino libre medio l— dependerá de cuántas moléculas hay alrededor y del "tamaño" de las moléculas, es decir, de lo grande que es el blanco que representan. El "tamaño" efectivo de un blanco en una colisión se describe habitualmente con una "sección transversal de colisión", el mismo concepto que se usa en la física nuclear o en problemas de dispersión de luz.

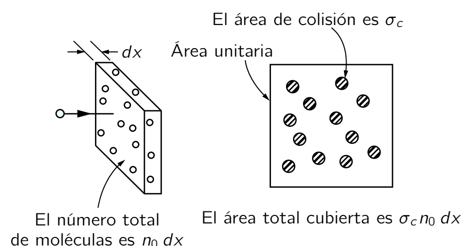


Figura 43-1. Sección eficaz de colisión.

Consideremos una partícula en movimiento que recorre una distancia dx a través de un gas que tiene n_0 dispersores (moléculas) por unidad de volumen (figura 43-1). Si observamos cada unidad de área perpendicular a la dirección de movimiento de la partícula que hemos seleccionado, encontraremos que hay n_0dx moléculas. Si cada una presenta un área efectiva de colisión, o como se llama comúnmente, "sección eficaz de colisión" σ_c, el área total cubierta por los dispersores es $\sigma_c n_0 dx$.

Por "sección eficaz de colisión" entendemos el área dentro de la cual debe estar ubicado el centro de nuestra partícula para que choque con una molécula determinada. Si las moléculas fuesen pequeñas esferas (representación clásica) esperaríamos que $\sigma_c = \pi(r_1+r_2)^2$, donde r_1 y r_2 son los radios de los dos objetos que chocan. La probabilidad de que nuestra partícula sufra una colisión es el cociente entre el área cubierta por las moléculas dispersoras y el área total, que hemos tomado igual a uno. Entonces, la probabilidad de una colisión al recorrer una distancia dx es simplemente $\sigma_c n_0 dx$:

$$\text{Probabilidad de una colisión en } dx = \sigma_c n_0\, dx. \tag{43.10}$$

Hemos visto más arriba que la probabilidad de una colisión en dx también se puede escribir en función del camino libre medio l como dx/l. Comparando esto con (43.10), podemos relacionar el camino libre medio con la sección eficaz de colisión:

$$\frac{1}{l} = \sigma_c n_0, \tag{43.11}$$

que es más fácil de recordar si la escribimos en la forma

$$\sigma_c n_0 l = 1. \tag{43.12}$$

Se puede interpretar que esta fórmula dice que debe haber una colisión, en promedio, cuando la partícula recorre una distancia l tal que las moléculas dispersoras *podrían* cubrir exactamente el área total. En un volumen cilíndrico de altura l y cuya área de la base es unitaria, hay $n_0 l$ dispersores, si cada uno tiene un área σ_c del área total cubierta, entonces $n_0 l\sigma_c$ es precisamente una unidad de área. Toda el área *no* está cubierta naturalmente, porque determinadas moléculas están parcialmente escondidas detrás de otras. Así, algunas moléculas van más allá de l antes de tener una colisión. Es sólo *en promedio* que las moléculas tienen una colisión cada vez que recorren la distancia l. Podemos determinar la sección eficaz de colisión σ_c a partir de medidas del camino libre medio l, y comparar el resultado con cálculos basados en una teoría detallada de la estructura atómica. ¡Pero esto ya es otro tema! Volvamos, pues, al problema de los estados fuera del equilibrio.

43-3 La velocidad de arrastre

Queremos describir lo que le sucede a una molécula, o a varias moléculas, que en algún aspecto son diferentes de la gran mayoría de las moléculas de un gas. Nos referiremos a las moléculas "de la mayoría" como moléculas "de fondo" y llamaremos moléculas "especiales", o moléculas S, para abreviar, a las moléculas

diferentes de las moléculas de fondo. Una molécula es posible que sea especial por cualquier número de razones: podría ser más pesada que las moléculas de fondo, podría ser un compuesto químico diferente, podría tener carga eléctrica —es decir, es posible que sea un ion en un fondo de moléculas no cargadas—. A causa de sus masas o cargas diferentes, las moléculas S pueden estar sujetas a fuerzas diferentes de las que actúan sobre las moléculas de fondo. Considerando lo que les sucede a estas moléculas S podemos comprender los efectos fundamentales que entran en juego de manera similar en muchos fenómenos diferentes. Para nombrar algunos: difusión de gases, corrientes eléctricas en baterías, sedimentación, separación centífruga, etcétera.

Comencemos concentrándonos en el proceso fundamental: una molécula S en un gas de fondo está sujeta a alguna fuerza especificable $\boldsymbol{F}$ (que podría ser, por ejemplo, gravitacional o eléctrica) y *además* a otras fuerzas no tan especificables debidas a las colisiones con las moléculas de fondo. Querríamos describir el comportamiento *general* de la molécula S. Lo que le ocurre, *en detalle*, es que se precipita de un lado para otro a medida que choca continuamente con otras moléculas. Pero si la observamos atentamente, vemos que hay cierto avance en la dirección de la fuerza $\boldsymbol{F}$. Decimos que hay un *arrastre* superpuesto a su movimiento al azar. Además, queremos saber cuál es la velocidad de su arrastre —*su velocidad de arrastre*— debida a la fuerza $\boldsymbol{F}$.

Si comenzamos a observar una molécula S en cualquier instante, podemos esperar que esté en alguna parte entre dos colisiones. Además de la velocidad con que quedó después de su última colisión, está adquiriendo una cierta componente de velocidad debido a la fuerza $\boldsymbol{F}$. En un corto tiempo (un tiempo τ en promedio) experimentará una colisión y empezará un nuevo tramo de su trayectoria. Tendrá una nueva velocidad inicial pero la misma aceleración debida a $\boldsymbol{F}$.

Para no complicar las cosas por el momento, supondremos que después de cada colisión nuestra molécula S tiene un comienzo completamente "nuevo". Es decir que no conserva el recuerdo de la aceleración pasada debida a $\boldsymbol{F}$. Ésta podría ser una hipótesis razonable si nuestra molécula S fuera mucho más liviana que las moléculas de fondo, pero seguramente no es válida en general. Más adelante discutiremos una hipótesis mejorada.

Por el momento, entonces, nuestra hipótesis es que la molécula S sale de cada colisión con una velocidad que puede estar en cualquier dirección con igual probabilidad. La velocidad inicial la llevará igualmente en todas direcciones y no contribuirá a ningún movimiento neto, por lo que no nos preocuparemos más de su velocidad inicial después de una colisión. Además de su movimiento al azar, cada molécula S tendrá, en cualquier momento, una velocidad adicional en la dirección de la fuerza $\boldsymbol{F}$, que ha adquirido *desde* su última colisión. ¿Cuál es el valor *medio* de *esta* parte de la velocidad? Es simplemente la aceleración $\boldsymbol{F}/m$ (donde m es la masa de la molécula S) multiplicada por el tiempo *medio* transcurrido *desde* la última colisión. Ahora bien, el tiempo medio *desde* la *última* colisión debe ser igual al tiempo medio *hasta* la colisión *siguiente*, al cual hemos llamado τ más arriba. Por supuesto, la velocidad *media* debida a $\boldsymbol{F}$ es precisamente lo que hemos llamado velocidad de arrastre, por lo que tenemos la relación:

$$v_{\text{arrastre}} = \frac{F\tau}{m}. \tag{43.13}$$

Esta relación básica es el nudo de nuestro argumento. Puede haber alguna complicación para determinar τ, pero el proceso fundamental está definido por la ecuación (43.13).

Notarán ustedes que la velocidad de arrastre es *proporcional* a la fuerza. Desafortunadamente no hay ningún nombre de uso general para la constante de proporcionalidad. Se han usado diferentes nombres para cada uno de los diversos tipos de fuerza. Si en un problema de electricidad la fuerza se escribe como la carga por el campo eléctrico, $\boldsymbol{F} = q\boldsymbol{E}$, la constante de proporcionalidad entre la velocidad y el campo eléctrico $\boldsymbol{E}$ se denomina comúnmente "movilidad". A pesar de la posibilidad de que haya un poco de confusión, utilizaremos el

término *movilidad* para el cociente entre la velocidad de arrastre y la fuerza, para *cualquiera* que sea ésta. Escribamos:

$$v_{\text{arrastre}} = \mu F. \tag{43.14}$$

en general, llamando movilidad a μ. Según la ecuación (43.13) tenemos

$$\mu = \tau/m. \tag{43.15}$$

La movilidad es proporcional al tiempo medio entre colisiones (hay menos colisiones para frenarla) e inversamente proporcional a la masa (mayor inercia significa menor velocidad adquirida entre colisiones).

Para obtener el coeficiente numérico correcto en la ecuación (43.13), que está correcta como está, se necesita algún cuidado. Sin intención de confundir, debemos aún señalar que los razonamientos tienen una sutileza que sólo se puede apreciar por medio de un estudio cuidadoso y detallado. Para ilustrar que hay dificultades a pesar de las apariencias, repetiremos el camino que nos llevó a la ecuación (43.13) de un modo razonable *pero errado* (¡y es el que se encontrará en muchos libros de texto!).

Podríamos haber dicho: el tiempo medio entre colisiones es τ. Después de una colisión la partícula parte con una velocidad al azar, pero adquiere una velocidad adicional entre colisiones que es igual a la aceleración por el tiempo. Como tarda el tiempo τ en llegar a la colisión *siguiente*, llega con la velocidad $(F/m)\tau$. En el momento de la colisión tenía velocidad nula. Luego entre las dos colisiones tiene en promedio una velocidad que es la mitad de la velocidad final, por lo que la velocidad media de arrastre es $\frac{1}{2}F\tau/m$. (¡Errado!) Este resultado es incorrecto y el de la ecuación (43.13) es correcto, aunque los argumentos puedan parecer igualmente satisfechos. La razón de que el segundo resultado esté equivocado es algo sutil y tiene que ver con lo siguiente: el razonamiento se hace como si todas las colisiones estuviesen separadas por el intervalo medio τ. El hecho es que algunos tiempos son más cortos y otros más largos que el intervalo medio. Los tiempos cortos se presentan *más a menudo* pero dan una contribución *menor* a la velocidad de arrastre porque dan menos oportunidad de “ponerse a andar realmente”. Si se toma debida cuenta de la *distribución* de tiempos libres entre colisiones, se puede demostrar que no debe estar el factor $\frac{1}{2}$, obtenido en el segundo razonamiento. El error se cometió al tratar de relacionar mediante un argumento simple la velocidad *final media* con la velocidad media misma. La relación no es simple, de modo que es mejor que nos concentremos en lo que queremos: la velocidad media misma. El primer razonamiento que hicimos determina la velocidad media directamente —¡y correctamente!—. ¡Pero quizás ahora podemos ver por qué en general no trataremos de obtener los coeficientes numéricos correctos en nuestras derivaciones elementales!

Volvamos ahora a nuestra hipótesis simplificada: cada colisión suprime el recuerdo del movimiento pasado —después de cada colisión se tiene una nueva partida—. Supongan que nuestra molécula S es un objeto pesado en un fondo de moléculas más ligeras. Por lo tanto, nuestra molécula S no perderá su momentum “hacia adelante” en cada colisión. Serían necesarias varias colisiones para que su movimiento fuera “azaroso” de nuevo. En su lugar deberíamos suponer que en cada colisión —en cada intervalo τ en promedio— pierde cierta fracción de su momentum. No elaboraremos los detalles, sino que diremos simplemente que el resultado es equivalente a remplazar τ, tiempo medio de colisión, por un τ nuevo —y más largo que corresponde al “tiempo medio de olvido”, el tiempo medio que tarda en olvidar su momentum hacia adelante—. Con esta interpretación de τ podemos usar nuestra fórmula (43.15) en situaciones que no son tan simples como la que supusimos inicialmente.

43-4 Conductividad iónica

Apliquemos ahora nuestros resultados a un caso particular. Supongan que tenemos un gas en un recipiente en el que también hay algunos iones —átomos

o moléculas con una carga eléctrica neta—. Mostramos esquemáticamente la situación en la figura 43-2. Si dos paredes opuestas del recipiente son placas metálicas, podemos conectarlas a los terminales de una batería y producir así un campo eléctrico en el gas. El campo eléctrico dará lugar a una fuerza sobre los iones, por lo que comenzarán a desplazarse hacia una u otra de las placas. Se inducirá una corriente eléctrica y el gas con sus iones se comportará como un resistor. Calculando el flujo de iones a partir de la velocidad de arrastre podemos determinar la resistencia. Preguntamos, específicamente, ¿cómo depende el flujo de corriente eléctrica de la diferencia de potencial V que aplicamos a las dos placas?

Consideremos el caso de que nuestro recipiente es una caja rectangular de longitud b y sección transversal A (figura 43-2). Si la diferencia de potencial, o voltaje, entre una placa y otra es V, el campo eléctrico E entre las placas es V/b. (El potencial eléctrico es el trabajo que se hace al llevar una carga unitaria de una placa a la otra. La fuerza sobre una carga unitaria es $\boldsymbol{E}$. Si $\boldsymbol{E}$ es el mismo en todos los puntos entre las placas, lo cual es una aproximación suficientemente buena por ahora, el trabajo realizado sobre una carga unitaria es simplemente Eb, de donde $V = Eb$.) La fuerza especial sobre un ion del gas es $q\boldsymbol{E}$, de donde q es la carga del ion. La velocidad de arrastre del ion es entonces μ por esta fuerza, podemos decir que es

$$v_{\text{arrastre}} = \mu F = \mu q E = \mu q \frac{V}{b}. \tag{43.16}$$

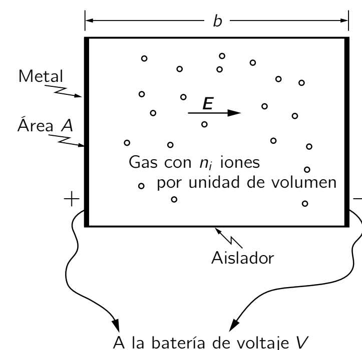


Figura 43-2. Corriente eléctrica en un gas ionizado.

Una corriente eléctrica I es el flujo de carga en la unidad de tiempo. La corriente eléctrica hacia una de las placas está dada por la carga total de los iones que lleguen a la placa en la unidad de tiempo. Si los iones se desplazan hacia la placa con la velocidad v_{arrastre}, los que están dentro de una distancia $(v_{\text{arrastre}} \cdot T)$ llegarán a la placa en el tiempo T. Si hay n_i iones por unidad de volumen, el número que llega a la placa en el tiempo T es $(n_i \cdot A \cdot v_{\text{arrastre}} \cdot T)$. Cada ion lleva una carga q, por lo que tenemos:

$$\text{Carga recogida durante } T = q n_i A v_{\text{arrastre}} T. \tag{43.17}$$

La corriente I es la carga recogida durante T dividida entre T, de donde

$$I = q n_i A v_{\text{arrastre}}. \tag{43.18}$$

Sustituyendo v_{arrastre} dada por (43.16), tenemos:

$$I = \mu q^2 n_i \frac{A}{b} V. \tag{43.19}$$

Encontramos que la corriente es proporcional al voltaje, lo cual es precisamente la forma, de la ley de Ohm, y que la resistencia R es la inversa de la constante de proporcionalidad:

$$\frac{1}{R} = \mu q^2 n_i \frac{A}{b}. \tag{43.20}$$

Tenemos una relación entre la resistencia y las propiedades moleculares n_i, q y μ, que a su vez depende de m y de τ. Si conocemos n_i y q por mediciones atómicas, se podría usar una medida de R para determinar μ, y de μ también τ.

43-5 Difusión molecular

Pasamos ahora a un problema y a un análisis de diferente tipo, la teoría de la difusión. Supongan que tenemos un recipiente con gas en equilibrio térmico y que en algún lugar del recipiente introducimos una pequeña cantidad de gas de un tipo diferente. Llamaremos gas "de fondo" al gas original, y gas "especial" al nuevo. El gas especial comenzará a extenderse por todo el recipiente, pero lo hará lentamente debido a la presencia del gas de fondo. Este lento proceso de extensión se llama *difusión*. La difusión está controlada principalmente por

los golpes que las moléculas del gas especial reciben de las moléculas del gas de fondo. Después de un gran número de colisiones, las moléculas especiales acaban extendidas más o menos uniformemente por todo el volumen. Debemos tener cuidado de *no* confundir la difusión de un gas con el transporte macroscópico que puede ocurrir debido a corrientes de convección. Lo más común es que la mezcla de dos gases ocurra mediante una combinación de convección y difusión. Ahora sólo estamos interesados en el caso en que *no* hay *corrientes "de viento"*. El gas se extiende únicamente por los movimientos moleculares, por difusión. Queremos calcular a qué velocidad tiene lugar la difusión.

Calculemos ahora el *flujo neto* de moléculas del gas "especial" debido a los movimientos moleculares. Sólo habrá un flujo neto cuando hay una distribución no uniforme de las moléculas, de lo contrario todos los movimientos moleculares se promediarían sin dar ningún flujo neto. Consideremos primero el flujo en la dirección x. Para hallarlo consideremos un plano imaginario perpendicular al eje x y contemos el número de moléculas especiales que atraviesan este plano. Para obtener el flujo neto, debemos contar como positivas las moléculas que cruzan en la dirección positiva de x y *restar* de este número el de las que cruzan en la dirección x negativa. Como hemos visto muchas veces, el número de las que atraviesan el área de una superficie en un tiempo ΔT está dado por el número de las que al comienzo del intervalo ΔT están en un volumen que se extiende a una distancia $v\Delta T$ del plano. (Observen que aquí v es la verdadera velocidad molecular y no la velocidad de arrastre.)

Simplificaremos nuestra álgebra dando un área unitaria a nuestra superficie. El número de moléculas especiales que pasan de izquierda a derecha (tomando la dirección $+x$ hacia la derecha) es $n_-v\Delta T$, donde n_- es el número de moléculas especiales que hay por unidad de volumen a la izquierda (a menos de un factor 2 o algo así, pero estamos ignorando esos factores). Análogamente, el número que cruza de derecha a izquierda es $n_+v\Delta T$, donde n_+ es la densidad numérica de moléculas especiales a la derecha del plano. Si llamamos J a la corriente molecular, con lo cual entendemos el flujo neto de moléculas por unidad de área y por unidad de tiempo, tenemos:

$$J = \frac{n_- v\,\Delta T - n_+ v\,\Delta T}{\Delta T}, \tag{43.21}$$

o sea

$$J = (n_- - n_+)v. \tag{43.22}$$

¿Qué usar para n_- y n_+? Cuando decimos "la densidad a la izquierda", ¿hasta *qué distancia* a la izquierda entendemos? Deberíamos elegir la densidad en el lugar desde donde las moléculas empezaron su "vuelo", porque el número de las que *empiezan* esos viajes está determinado por el número de las que están presentes en ese lugar. En consecuencia, debemos entender por n_- la densidad a una distancia a la izquierda igual al camino libre medio l, y por n_+ la densidad a una distancia l a la derecha de nuestra superficie imaginaria.

Es conveniente considerar que la distribución espacial de nuestras moléculas especiales está descrita por una función continua de x, y y z que llamaremos n_a. Por $n_a(x,y,z)$ entendemos la densidad numérica de moléculas especiales en un pequeño elemento de volumen centrado en (x,y,z). En términos de n_a, podemos expresar la diferencia $(n_+ - n_-)$ en la forma:

$$(n_+ - n_-) = \frac{dn_a}{dx}\,\Delta x = \frac{dn_a}{dx}\cdot 2l. \tag{43.23}$$

Sustituyendo este resultado en la ecuación (43.22) y despreciando el factor 2 de ésta, obtenemos:

$$J_x = -lv\,\frac{dn_a}{dx}. \tag{43.24}$$

Hemos encontrado que el flujo de moléculas especiales es proporcional a la derivada de la densidad, a lo que a veces se denomina "gradiente" de la densidad.

Está claro que hemos hecho varias operaciones aproximadas. Además, hemos dejado de lado varios factores como el dos de la derecha de la ecuación, hemos usado v donde deberíamos haber usado v_x y hemos supuesto que n_+ y n_- se refieren a puntos a una distancia perpendicular l de nuestra superficie, mientras que para las moléculas que no viajan perpendicularmente al elemento de superficie, l debería corresponder a la distancia *oblicua* desde la superficie. Es posible hacer todos estos refinamientos; el resultado de un análisis más cuidadoso muestra que el segundo miembro de la ecuación (43.24) se debe multiplicar por 1/3. Luego, una respuesta mejor es:

$$J_x = -\frac{lv}{3}\,\frac{dn_a}{dx}. \tag{43.25}$$

Podemos escribir ecuaciones similares para las corrientes en las direcciones y y z.

La corriente J_x y el gradiente de densidad dn_a/dx se pueden medir por medio de observaciones macroscópicas. Su cociente determinado experimentalmente se denomina "coeficiente de difusión" D. Esto es,

$$J_x = -D\,\frac{dn_a}{dx}. \tag{43.26}$$

Hemos podido demostrar que para un gas es de esperar que

$$D = \tfrac{1}{3}lv. \tag{43.27}$$

Hasta aquí, en este capítulo, hemos considerado dos procesos distintos: *movilidad*, o sea el arrastre de moléculas debido a fuerzas "de afuera", y *difusión*, o sea el extenderse determinado únicamente por las fuerzas internas, las colisiones al azar. Hay sin embargo, una relación entre ellos, ya que ambos dependen básicamente de los movimientos térmicos, y el camino libre medio l aparece en ambos cálculos.

Si en la ecuación (43.25) sustituimos $l = v\tau$ y $\tau = \mu m$, tenemos:

$$J_x = -\tfrac{1}{3}mv^2\mu\,\frac{dn_a}{dx}. \tag{43.28}$$

Pero mv^2 depende únicamente de la temperatura. Recordemos que:

$$\tfrac{1}{2}mv^2 = \tfrac{3}{2}kT, \tag{43.29}$$

por lo que

$$J_x = -\mu kT\,\frac{dn_a}{dx}. \tag{43.30}$$

Encontramos que el coeficiente de *difusión* D es sencillamente kT por el coeficiente de *movilidad* μ:

$$D = \mu kT. \tag{43.31}$$

Y resulta que el coeficiente numérico en (43.31) es exactamente correcto —no es necesario introducir factores adicionales para compensar nuestras operaciones aproximadas—. Podemos demostrar, en verdad, que (43.31) *siempre* debe ser correcta —aun en situaciones complicadas (por ejemplo, el caso de una suspensión en un líquido) donde los detalles de nuestro cálculo simple no servirían de ninguna manera—.

Para demostrar que (43.31) debe ser correcta en general, la deduciremos de un modo diferente, empleando únicamente nuestros principios básicos de mecánica estadística. Imaginen una situación en que hay un gradiente de moléculas "especiales" y tenemos una corriente de difusión proporcional al gradiente de densidad, conforme a la ecuación (43.26). Aplicamos ahora un campo de fuerzas en la dirección x de modo que cada molécula especial sienta la fuerza F. Conforme a la *definición* de movilidad μ habrá una velocidad de arrastre dada por

$$v_{\text{arrastre}} = \mu F. \tag{43.32}$$

Según nuestro razonamiento habitual, la *corriente de arrastre* (número *neto* de moléculas que atraviesan la unidad de área en la unidad de tiempo) será:

$$J_{\text{arrastre}} = n_a v_{\text{arrastre}}, \tag{43.33}$$

o sea

$$J_{\text{arrastre}} = n_a \mu F. \tag{43.34}$$

Ahora *ajustamos* la fuerza F de modo que la corriente de arrastre debida a F *compense* exactamente la difusión y *no haya flujo neto* de nuestras moléculas especiales. Tenemos $J_x + J_{\text{arrastre}} = 0$, o sea

$$D\,\frac{dn_a}{dx} = n_a \mu F. \tag{43.35}$$

En condiciones "de compensación" encontramos un gradiente de densidad fijo (en el tiempo) dado por

$$\frac{dn_a}{dx} = \frac{n_a \mu F}{D}. \tag{43.36}$$

¡Pero vean ustedes! Estamos describiendo un estado de *equilibrio* por lo que sirven las leyes de *equilibrio* de la mecánica estadística. Según estas leyes la probabilidad de encontrar una molécula en la coordenada x es proporcional a $e^{-U/kT}$, donde U es la energía potencial. En función de la densidad numérica n_a, esto significa que:

$$n_a = n_0 e^{-U/kT}. \tag{43.37}$$

Si derivamos (43.37) respecto a x, encontramos:

$$\frac{dn_a}{dx} = -n_0 e^{-U/kT} \cdot \frac{1}{kT}\,\frac{dU}{dx}, \tag{43.38}$$

o sea

$$\frac{dn_a}{dx} = -\frac{n_a}{kT}\,\frac{dU}{dx}. \tag{43.39}$$

En nuestra situación, como la fuerza F está en la dirección x, la energía potencial U es simplemente $-F_x$, y $-dU/dx = F$. La ecuación (43.39) da entonces

$$\frac{dn_a}{dx} = \frac{n_a F}{kT}. \tag{43.40}$$

[Ésta es exactamente la ecuación (40.2), de la cual dedujimos $e^{-U/kT}$ inicialmente, por lo que volvemos al punto de partida.] Comparando (43.40) con (43.36) obtenemos exactamente la ecuación (43.31). Hemos demostrado que la ecuación (43.31), que da la corriente de difusión en función de la movilidad, tiene el coeficiente correcto y es válida en forma muy general. La movilidad y la difusión están íntimamente relacionadas. Einstein fue el primero en deducir esta relación.

43-6 Conductividad térmica

Los métodos de la teoría cinética que hemos estado empleando más arriba también se pueden emplear para calcular la *conductividad térmica* de un gas. Si el gas que está en la parte de encima del recipiente está más caliente que el de abajo, habrá un flujo de calor desde arriba hacia abajo. (Consideramos que la parte de arriba está más caliente porque de otra manera se establecerían corrientes de convección y el problema ya no sería de *conducción* del calor.) Así que, la transferencia de calor desde el gas más caliente hasta el más frío se debe a la difusión de las moléculas "calientes" —las de mayor energía— hacia abajo y a la difusión de las moléculas "frías" hacia arriba. Para calcular el flujo de energía térmica podemos buscar la energía transportada hacia abajo a través de un elemento de superficie por las moléculas que se mueven hacia abajo, y la

energía transportada hacia arriba a través de dicho elemento por las moléculas que se mueven hacia arriba. La diferencia nos dará el flujo neto de energía hacia abajo.

Se define la conductividad térmica κ como el cociente entre la rapidez con que se transporta energía térmica a través de un área unitaria, y el gradiente de temperatura:

$$\frac{1}{A}\frac{dQ}{dt} = -\kappa\,\frac{dT}{dz}. \tag{43.41}$$

Como los detalles de los cálculos son muy similares a los que hemos realizado antes al considerar difusión molecular, dejaremos al lector como ejercicio demostrar que:

$$\kappa = \frac{knlv}{\gamma - 1}, \tag{43.42}$$

donde $kT/(\gamma - 1)$ es la energía media de una molécula a la temperatura T. Si usamos nuestra relación $nl\sigma_c = 1$, se puede escribir la conductividad calórica en la forma

$$\kappa = \frac{1}{\gamma - 1}\,\frac{kv}{\sigma_c}. \tag{43.43}$$

Tenemos un resultado bastante sorprendente. Sabemos que la velocidad media de las moléculas de gas depende de la temperatura pero *no de la densidad*. Es de esperar que σ_c sólo dependa del *tamaño* de las moléculas. Por lo tanto, nuestro simple resultado dice que la conductividad térmica κ (y por lo tanto, la *rapidez* de flujo de calor en cualquier circunstancia particular) ¡es independiente de la *densidad* del gas! La variación del número de "portadores" de energía, con una variación de la densidad, está compensada por la distancia mayor que los "portadores" pueden andar entre colisiones.

Se podría preguntar: "¿es el flujo de calor independiente de la densidad del gas en el límite de la densidad tiende a cero? ¿Cuándo no hay nada de gas?" ¡Por cierto que no! La fórmula (43.43) fue obtenida, como todas las otras de este capítulo, con la hipótesis de que el camino libre medio entre colisiones es mucho menor que cualquiera de las dimensiones del recipiente. En cuanto la densidad del gas es tan baja que una molécula tiene una buena probabilidad de cruzar de una pared a otra del recipiente sin tener una colisión, dejan de ser válidos todos los cálculos hechos en este capítulo. En esos casos debemos volver a la teoría cinética y calcular de nuevo los detalles de lo que ocurrirá.

44

Leyes de la termodinámica

44-1 Máquinas térmicas, primera ley

Hasta ahora hemos estado discutiendo las propiedades de la materia desde un punto de vista atómico, tratando de entender más o menos lo que sucederá si suponemos que las cosas están formadas de átomos que obedecen ciertas leyes. Hay, sin embargo, un número de relaciones entre las propiedades de las sustancias que se pueden calcular sin tener en cuenta una estructura detallada de los materiales. La determinación de las relaciones entre las diversas propiedades de los materiales, sin tener en cuenta su estructura interna, es el objeto de la *termodinámica*. La termodinámica se desarrolló, históricamente, antes de que se hubiera alcanzado una comprensión de la estructura interna de la materia.

Daremos un ejemplo: sabemos por la teoría cinética que la presión de un gas es causada por el bombardeo molecular y sabemos que si calentamos un gas, para que así aumente el bombardeo, la presión debe aumentar. Inversamente, si el pistón de un recipiente con gas se mueve hacia adentro contra la fuerza del bombardeo, aumentará la energía de las moléculas que bombardeen el pistón y consecuentemente aumentará la temperatura. Así, pues, por un lado, si aumentamos la temperatura a volumen constante, incrementamos la presión. Y por otro lado, si comprimimos el gas, encontraremos que la temperatura aumentará. Mediante la teoría cinética, se puede derivar una relación cuantitativa entre estos dos efectos, pero se podría presumir instintivamente que se relacionan en alguna forma necesaria que es independiente a los pormenores de las colisiones.

Consideremos otro ejemplo. Mucha gente está familiarizada con esta propiedad interesante de la goma: si tomamos una banda elástica y la estiramos, se calienta. Si alguien se coloca entre los labios, por ejemplo, y la estira puede sentir claramente un calentamiento y éste es reversible en el sentido de que si suelta la banda rápidamente mientras la tiene en los labios, nota claro un enfriamiento. Esto significa que cuando estiramos una banda de goma se calienta y que cuando le quitamos la tensión se enfría. Ahora bien, nuestro instinto nos puede sugerir que si calentásemos una banda, ésta podría encogerse: el hecho de que al estirar una banda se caliente podría implicar que el calentarla causase su contracción. Y en realidad, si aplicamos una llama de gas a una banda elástica que sostiene un peso veremos que la banda se contrae abruptamente (figura 44-1). Por consiguiente, es verdad que cuando calentamos una banda elástica se encoge, y este hecho está ciertamente relacionado con el otro de que cuando le suprimimos la tensión, se enfría.

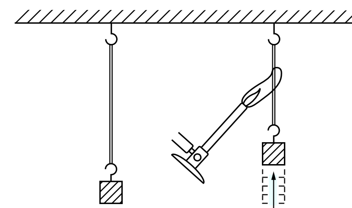


Figura 44-1. La banda elástica calentada.

La maquinaria interna de la goma causante de estos efectos es bastante complicada. La describiremos con alguna extensión desde un punto de vista molecular, aunque nuestra principal intención en este capítulo sea entender la relación de estos efectos independientemente del modelo molecular. A pesar de ello, podemos mostrarles a partir del modelo molecular que estos efectos están íntimamente relacionados. Un modo de entender el comportamiento de la goma es reconocer que esta sustancia consiste en un gran enredo de cadenas largas de moléculas, una especie de "espagueti molecular", con una complicación adicional: entre las cadenas hay enlaces —como un espagueti que algunas veces se suelda con otro cuando lo cruza—, magnífico enredo. Cuando tiramos de tal enredo, algunas de las cadenas tienden a alinearse en la dirección del tirón. Al mismo tiempo, estando las cadenas en movimiento térmico chocan continuamente una con otra. De lo que se sigue es que si se ha alargado una cadena no permanecerá por sí misma alargada, ya que las otras cadenas y moléculas la golpearán lateralmente

y tenderá a encogerse de nuevo. Por lo cual, la verdadera razón por la que una banda elástica tiende a contraerse es ésta: cuando se tira de ella, las cadenas se alargan y las agitaciones térmicas de las moléculas circundantes tienden a contraerlas y hacer que se acorten. Se puede apreciar que si se mantienen las cadenas estiradas y se aumenta la temperatura, con lo que también se aumenta la intensidad del bombardeo alrededor de las cadenas, éstas tienden a encogerse y son capaces de arrastrar un peso más grande cuando se las calienta. Si se le permite a una banda elástica que se relaje, después de haber estado estirada por algún tiempo, cada cadena se suaviza y las moléculas que la golpean pierden energía a medida que golpean la cadena que se relaja. Así, la temperatura cae.

Hemos visto cómo se pueden relacionar estos dos procesos, contracción por calentamiento y enfriamiento durante la relajación, mediante la teoría cinética, pero sería un tremendo desafío para ella determinar la relación precisa entre los dos. Tendríamos que saber cuántas colisiones ocurrirían por segundo y cómo son las cadenas, y tendríamos que tener en cuenta otras complicaciones de toda clase. El mecanismo detallado es tan complejo que no podemos en realidad determinar exactamente lo que sucede mediante la teoría cinética; aun así, ¡podemos calcular una relación determinada entre los dos efectos que observamos sin conocer nada de la maquinaria interna!

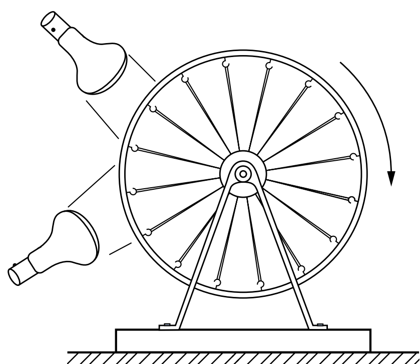


Figura 44-2. La máquina térmica de banda elástica.

Todo el objeto de la termodinámica depende esencialmente de la siguiente consideración: ya que una banda elástica es “más fuerte” a altas que a bajas temperaturas, debería ser posible levantar pesos y hacerlos girar y, por lo tanto, realizar un trabajo con calor. De hecho, hemos visto ya experimentalmente que una banda calentada puede levantar un peso. El estudio de cómo se realiza trabajo con calor es el comienzo de la ciencia de la termodinámica. ¿Podemos hacer que una máquina que utiliza el efecto de calentamiento de una banda realice trabajo? Se puede construir una máquina que parece tonta y que realiza esto exactamente. Consiste en una rueda de bicicleta en la que todos los radios son bandas elásticas (figura 44-2). Si calentamos las bandas de un lado de la rueda con un par de lámparas, las bandas se hacen “más fuertes” que las del otro lado. El centro de gravedad de la rueda se desviará a un lado, apartándose del eje, por lo que la rueda girará. A medida que gira, las bandas frías se mueven hacia el calor, y las calientes se alejan y se enfrían, y de ese modo la rueda gira despacio mientras tengamos aplicado calor. La eficiencia de esta máquina es extremadamente baja. Cuatrocientos vatios de potencia se meten en las lámparas, ¡y solamente es posible levantar una mosca con esa máquina! Una cuestión más interesante es si podemos o no obtener calor para realizar trabajo de maneras más eficientes.

La ciencia de la termodinámica comenzó, en realidad, con un análisis que el gran ingeniero Sadi Carnot hizo del problema: cómo construir la máquina mejor y más eficiente; y éste es uno de los pocos casos famosos en los que la ingeniería ha contribuido fundamentalmente a la teoría física. Otro ejemplo que nos viene a la memoria es el análisis más reciente de la teoría de la información por Claude Shannon. Entre paréntesis, resulta que estos dos análisis están íntimamente relacionados.

Ahora bien, el modo en que una máquina de vapor trabaja ordinariamente es que el calor de un fuego evapora agua y el vapor así formado se expande y empuja un pistón que hace que una rueda gire. Así que el vapor empuja el pistón —¿y entonces qué?—. Hay que terminar el trabajo: un modo tonto de completar el ciclo sería dejar que el vapor escapase al aire, pero entonces hay que estar suministrando agua. Es más barato —más eficiente— enviar el vapor a otra caja, donde se lo condensa mediante agua fría, y entonces bombear el agua de nuevo de la caldera de modo que circule continuamente. Se suministra, pues, calor a la máquina y se lo convierte en trabajo. Pero, ¿no sería mejor usar alcohol? ¿Qué propiedad debería tener una sustancia para que la máquina fuese la mejor posible? Ésta fue la pregunta que Carnot se hizo y uno de los resultados adicionales fue el descubrimiento del tipo de relación que hemos explicado antes.

Los resultados de la termodinámica están todos implícitamente contenidos en ciertas proposiciones, aparentemente simples, llamadas *leyes de la termodinámica*.

En la época en que vivió Carnot, la primera ley de la termodinámica, la conservación de la energía, no era conocida. Los razonamientos de Carnot, sin embargo, fueron hechos tan cuidadosamente que ¡son válidos aunque en su época no se conociese la primera ley! Algún tiempo después, Clapeyron hizo una derivación más sencilla y que se pudo entender más fácilmente que el razonamiento tan sutil de Carnot. Pero resultó que Clapeyron supuso, no la conservación de la energía en general, sino que el *calor* se conservaba de acuerdo a la teoría calórica, luego se demostró que era falsa. Por esto se ha dicho con frecuencia que el razonamiento de Carnot era erróneo. Pero su razonamiento era totalmente correcto. Solamente la versión simplificada de Clapeyron, que todo el mundo lee, era incorrecta.

Así, ¡la segunda ley de la termodinámica fue descubierta por Carnot antes que la primera ley! Sería interesante darles el razonamiento que hizo Carnot sin usar la primera ley, pero no lo haremos porque queremos aprender física y no historia. Usaremos la primera ley desde el principio, a pesar de que se puede hacer una gran cantidad de cosas sin ella.

Comencemos estableciendo la primera ley, la conservación de la energía: si se tiene un sistema y se lo calienta y se realiza trabajo sobre él, entonces su energía aumenta en el calor que se le ha dado y en el trabajo que se ha efectuado. Podemos escribir esto como sigue: el calor Q dado al sistema más el trabajo W efectuado sobre el sistema, es el aumento de energía U del mismo; esta energía se llama a veces energía interna:

$$\text{Variación de } U = Q + W. \tag{44.1}$$

La variación de U se puede representar como añadiendo un poco de calor ΔQ y añadiendo un poco de trabajo ΔW:

$$\Delta U = \Delta Q + \Delta W, \tag{44.2}$$

que es la forma diferencial de la misma ley. Sabemos esto muy bien por un capítulo anterior.

44-2 Segunda ley

Ahora, ¿y qué pasa con la segunda ley de la termodinámica? Sabemos que si realizamos un trabajo contra la fricción, digamos, el trabajo perdido es igual al calor producido. Si realizamos trabajo en una habitación a temperatura T y lo realizamos muy despacio, la temperatura de la habitación no cambia mucho y hemos convertido trabajo en calor a una temperatura dada. ¿Qué hay de la posibilidad inversa? ¿Es posible convertir calor en trabajo a una temperatura dada? La segunda ley de la termodinámica afirma que no es posible. Sería muy conveniente que pudiésemos convertir calor en trabajo simplemente invirtiendo un proceso como el de fricción. Si consideramos solamente la conservación de la energía, podenis pensar que la energía térmica, como la de los movimientos vibracionales de las moléculas, pueden dar un suministro bueno de energía útil. Pero Carnot supuso que es imposible extraer energía del calor a una sola temperatura. Es decir, si todo el mundo estuviera a la misma temperatura, su energía térmica no se podría convertir en trabajo: mientras que el proceso de convertir trabajo en calor puede ocurrir a una temperatura dada, no puede ocurrir el proceso inverso de que el calor produzca trabajo. Específicamente, Carnot supuso que no se puede tomar calor a una cierta temperatura y convertirlo en trabajo *sin ningún otro cambio* en el sistema o en el medio ambiente.

La última frase es muy importante. Supongan que tenemos una lata de aire comprimido a una cierta temperatura y dejamos que el aire se expanda. Puede realizar un trabajo, puede, por ejemplo, mover un martillo. Se enfría un poco en la expansión, pero si tuviésemos un gran mar, como el océano, a una temperatura dada —un depósito de calor—, lo podríamos calentar de nuevo. Por lo tanto, hemos sacado calor del mar y hemos realizado un trabajo con el aire comprimido. Pero Carnot no estaba equivocado, ya que *no lo hemos dejado todo como estaba*. Si volvemos a comprimir el aire que dejamos expandir, encontraremos que estamos

realizando un trabajo adicional, y cuando hayamos terminado descubriremos que no sólo no hemos obtenido ningún trabajo del sistema a la temperatura T, sino que realmente hemos puesto alguno dentro. Debemos hablar solamente de situaciones en las que el *resultado neto* de todo el proceso es sacar calor y convertirlo en trabajo, del mismo modo que el resultado neto de realizar un trabajo contra la fricción es tomar trabajo y convertirlo en calor. Si nos movemos en un camino cerrado, podemos llevar el sistema precisamente a su punto de partida, con el resultado neto de que realizamos trabajo contra la fricción y produjimos calor. ¿Podemos invertir el proceso? Accionen un interruptor para que todo vaya al revés, ¿es que la fricción realiza trabajo en contra nuestra y enfría el mar? Según Carnot: ¡no! Por consiguiente, supongamos que esto es imposible.

Si ello fuera posible, significaría, entre otras cosas, que podríamos sacar calor de un cuerpo frío y dárselo a un cuerpo caliente sin ningún gasto, por así decir. Pero nosotros sabemos que es natural que una cosa caliente dé calor a una fría, si colocamos simplemente juntos un cuerpo caliente y otro frío, y nada más, ¡nuestra experiencia nos asegura que no va a suceder que el cuerpo caliente se ponga más caliente y el frío más frío! Pero podríamos obtener trabajo extrayendo calor del océano, digamos, o de cualquier otra cosa a una sola temperatura y ese trabajo se podría convertir de nuevo en calor mediante fricción a otra temperatura. Por ejemplo, el otro brazo de una máquina que produce trabajo podría estar frotando algo que ya está caliente. El resultado neto sería sacar calor de un cuerpo "frío", el océano, y dárselo a un cuerpo caliente. Ahora bien, la hipótesis de Carnot, la segunda ley de la termodinámica, algunas veces se enuncia como sigue: el calor no puede, por sí mismo, fluir de un objeto frío a uno caliente. Pero, como acabamos de ver, estos dos enunciados son equivalentes: en primer lugar, que no se puede idear un proceso cuyo único resultado es convertir calor en trabajo a una sola temperatura, y en segundo lugar que no se puede hacer que el calor por sí mismo pase de un cuerpo frío a uno caliente. Usaremos principalmente la primera forma.

El análisis de Carnot de las máquinas térmicas es muy semejante al razonamiento que hicimos sobre las máquinas elevadoras de peso en nuestra discusión de la conservación de la energía en el capítulo 4. En realidad, dicho razonamiento tomó como modelo el de Carnot sobre las máquinas térmicas, por lo que el presente análisis les sonará muy parecido.

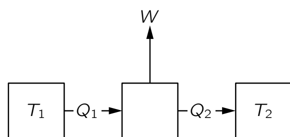


Figura 44-3. Máquina térmica.

Supongan que construimos una máquina térmica que contiene en algún lugar una "caldera" a una temperatura T_1. Se saca un cierto calor Q_1 de la caldera, la máquina de vapor realiza algún trabajo W, y también se entrega cierto calor Q_2 a un "condensador" a otra temperatura T_2 (figura 44-3). Carnot no dijo cuánto calor, ya que no conocía la primera ley, ni tampoco usó la ley de que Q_2 era igual a Q_1 porque no creía en ella. Aunque todo el mundo pensaba que de acuerdo con la teoría térmica los calores Q_1 y Q_2 tendrían que ser iguales, Carnot no lo dijo —esto es parte de lo habilidoso de su razonamiento—. Si nosotros sí usamos la primera ley, encontramos que el calor entregado Q_2 es igual al calor sacado Q_1 menos el trabajo realizado W:

$$Q_2 = Q_1 - W. \tag{44.3}$$

(Si tenemos una especie de proceso cíclico donde se devuelve el agua a la caldera después de haber sido condensada, diremos que hemos absorbido el calor Q_1 y hemos realizado el trabajo W, durante cada ciclo, para una cierta cantidad de agua que ha completado el ciclo.)

Ahora construiremos otra máquina y veremos si no podemos obtener más trabajo entregando la misma cantidad de calor a la temperatura T_1, con el condensador aún a la temperatura T_2. Usaremos la misma cantidad de calor Q_1 de la caldera y trataremos de obtener más trabajo del que obtuvimos de la máquina de vapor, quizás usando otro fluido, por ejemplo, alcohol.

44-3 Máquinas reversibles

Ahora debemos analizar nuestras máquinas. Hay una cosa clara: perderemos algo si las máquinas contienen partes en las que haya fricción. La mejor máquina será una sin fricción. Suponemos, entonces, la misma idealización que hicimos cuando estudiamos la conservación de la energía, esto es, una máquina sin fricción de ninguna clase.

Debemos considerar también lo análogo del movimiento sin fricción: transferencia de calor "sin fricción". Si colocamos un objeto caliente a una alta temperatura junto a uno frío, para que fluya el calor, entonces no es posible hacer que el calor fluya en dirección opuesta mediante un cambio muy pequeño en la temperatura de cualquiera de los dos objetos. Pero cuando tenemos una máquina prácticamente sin fricción, si la empujamos con una pequeña fuerza en una dirección, marcha en esa dirección, y si la empujamos con una pequeña fuerza en la otra dirección, marcha en esa otra dirección. Necesitamos encontrar lo análogo al movimiento sin fricción: transferencia de calor cuya dirección podemos invertir con sólo un cambio pequeño. Si la diferencia de temperaturas es finita, es imposible, pero si nos aseguramos de que el calor fluya siempre entre dos cosas esencialmente a la misma temperatura, con una diferencia infinitesimal solamente para que el flujo vaya en la dirección deseada, se dice que el flujo es reversible (véase figura 44-4). Si calentamos un poco el objeto de la izquierda, el calor fluirá hacia la derecha. Si lo enfriamos un poco, el calor fluirá hacia la izquierda. Por lo que encontramos que la máquina ideal es la llamada máquina *reversible*, en la que todo proceso es reversible en el sentido que mediante cambios mínimos, cambios infinitesimales, podemos hacer que la máquina vaya en la dirección opuesta. Esto significa que en ningún lugar de la máquina debe haber fricción apreciable, y que en ningún lugar de la misma debe haber un sitio donde el calor de los depósitos, o la llama de la caldera, esté en contacto directo con algo decididamente más frío o más caliente.

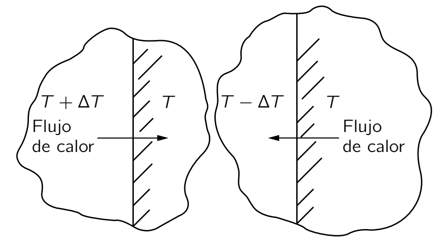


Figura 44-4. Transferencia reversible de calor.

Consideremos una máquina idealizada en la que todos los procesos son reversibles. Para mostrar que tal cosa es posible en principio, les daremos un ejemplo de un ciclo de máquinas que puede ser o no práctico, pero que por lo menos es reversible según la idea de Carnot. Supongan que tenemos un gas en un cilindro equipado con un pistón exento de fricción. El gas no es necesariamente un gas perfecto. El fluido ni siquiera tiene que ser un gas, pero para especificar digamos que tenemos un gas perfecto. Supongan, también, que tenemos dos fuentes de calor T_1 y T_2 —dos cosas enormes que tienen temperaturas fijas T_1 y T_2—. Supondremos en este caso que T_1 es mayor que T_2. Calentemos primero el gas dejando que se expanda al mismo tiempo, mientras está en contacto con la fuente de calor T_1. Mientras hacemos esto, tirando del pistón muy despacio cuando el calor fluye dentro del gas, nos aseguraremos de que la temperatura del gas será muy parecida a T_1. Si tiramos del pistón muy rápido, la temperatura del gas será mucho más baja que T_1 y así el proceso no será muy reversible, pero si tiramos de él lo suficientemente despacio, la temperatura del gas nunca se apartará mucho de T_1. Por otro lado, si empujamos el pistón hacia dentro muy despacio, la temperatura será solamente un infinitésimo más alta que T_1 y el calor será devuelto. Vemos que una expansión isotérmica de esta clase (temperatura constante), efectuada despacio y suavemente, es un proceso reversible.

Para entender lo que estamos haciendo, usaremos una representación (figura 44-6) de la presión del gas en función del volumen. Cuando el gas se expande la presión cae. La curva marcada (1) nos dice cómo varían la presión y el volumen si la temperatura se conserva fija en T_1. Para un gas ideal esta curva sería $PV = NkT_1$. Durante una expansión isotérmica, la presión disminuye cuando el volumen aumenta hasta que nos detenemos en el punto b. Al mismo tiempo, un cierto calor Q_1 debe fluir desde el depósito al gas, ya que si el gas se expandiese sin estar en contacto con el depósito, se enfriaría, como ya sabemos. Habiendo terminado la expansión isotérmica, deteniéndonos en el punto b, quitemos el cilindro del depósito y continuemos la expansión. Ahora no permitimos que entre calor en el cilindro. De nuevo realizamos la expansión despacio, por lo que no

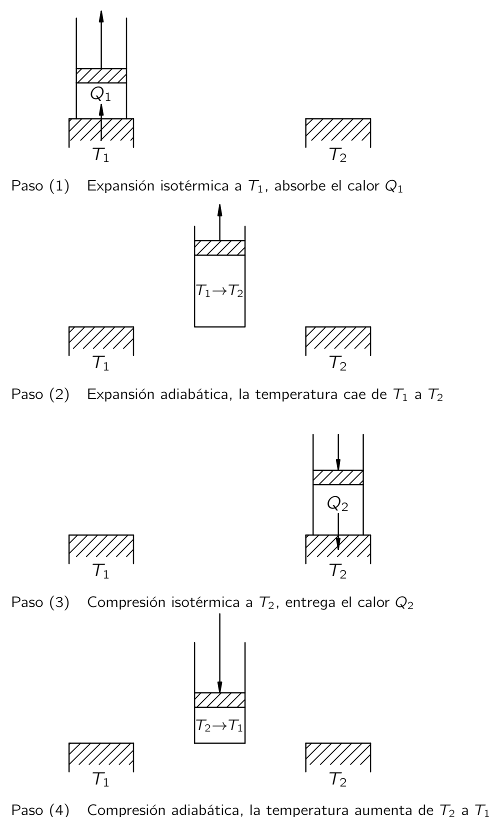


Figura 44-5. Pasos en el ciclo de Carnot.

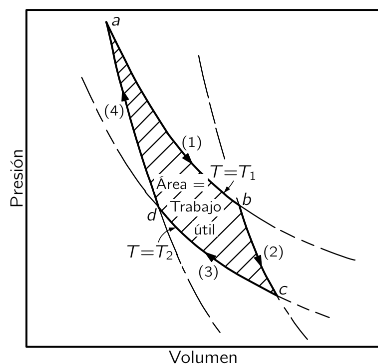


Figura 44-6. Ciclo de Carnot.

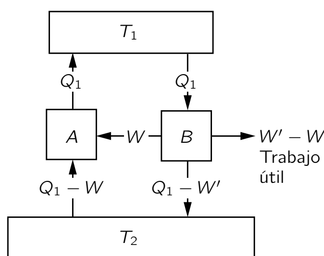


Figura 44-7. Máquina reversible A puesta a funcionar al revés por la máquina B.

hay razón para que no podamos invertirla y suponemos de nuevo que no hay fricción. El gas continúa expandiéndose y la temperatura disminuye, puesto que ahora no entra calor en el cilindro.

Dejamos que el gas se expanda, siguiendo la curva marcada (2), hasta que la temperatura baje a T_2 en el punto marcado c. Esta clase de expansión hecha sin añadir calor se llama expansión *adiabática*. Para un gas ideal ya sabemos que la curva (2) tiene la forma $PV^\gamma =$ constante, donde γ es una constante mayor que 1, de modo que la curva adiabática tiene una pendiente más negativa que la curva isotérmica. El cilindro de gas ha alcanzado ahora la temperatura T_2, por lo que si lo colocamos en la fuente de calor a la temperatura T_2 no habrá cambios irreversibles. Ahora comprimimos el gas muy despacio mientras está en contacto con el depósito a T_2, siguiendo la curva marcada (3) (figura 44-5, paso 3). Como el cilindro está en contacto con el depósito, la temperatura no aumenta, pero pasa calor Q_2 del cilindro al depósito a la temperatura T_2. Habiendo comprimido el gas isotérmicamente a lo largo de la curva (3) hasta el punto d, quitamos el cilindro de la fuente de calor a la temperatura T_2 y lo comprimimos aún más sin permitir que ningún calor salga. La temperatura aumentará y la presión seguirá la curva marcada (4). Si efectuamos bien cada paso, podemos volver al punto a a la temperatura T_1 donde comenzamos y repetir el ciclo.

Vemos que en este diagrama hemos llevado el gas alrededor de un ciclo completo, y durante el mismo le hemos dado Q_1 a la temperatura T_1 y le hemos quitado Q_1 a la temperatura T_2. Lo interesante es que este ciclo es reversible, por lo que podríamos representar todos los pasos al revés. Podríamos haber ido hacia atrás en lugar de hacia adelante: podríamos haber empezado en el punto a, a la temperatura T_1, expandir según la curva (4), expandir aún más hasta la temperatura T_2, absorbiendo calor Q_2, etc., haciendo el ciclo al revés. Si completamos el ciclo en un sentido, debemos aplicarle trabajo al gas, si lo hacemos en el otro sentido el gas nos devuelve el trabajo.

Entre paréntesis, es fácil calcular cuál es la cantidad total de trabajo, ya que el trabajo durante cualquier expansión es la presión por la variación de volumen, $\int P\,dV$. En este diagrama particular hemos representado P verticalmente y V horizontalmente. Por consiguiente, si llamamos y a la distancia vertical y x a la horizontal, nos da $\int y\,dx$ —en otras palabras, el área bajo la curva—. Por lo que el área bajo cada una de las curvas numeradas es una medida del trabajo efectuado por o sobre el gas en el paso correspondiente. No es difícil ver que el trabajo neto efectuado es el área sombreada de la figura.

Ahora que hemos dado un solo ejemplo de una máquina reversible, supondremos que son posibles otras máquinas de la misma especie. Supongamos que tenemos una máquina reversible A que toma Q_1 a T_1, hace el trabajo W y entrega algún calor a T_2. Supongamos ahora que tenemos cualquier otra máquina B, hecha por un hombre, ya diseñada o aún no inventada, formada de bandas elásticas, vapor o lo que sea, reversible o no, que está construida de tal forma que toma la misma cantidad de calor Q_1 a T_1 y lo cede a la temperatura inferior T_2 (figura 44-7). Suponemos que la máquina B efectúa algún trabajo W'. Demostraremos ahora que W' no es mayor que W —que ninguna máquina puede efectuar más trabajo que una reversible—. ¿Por qué? Supongan que en realidad W' fuese mayor que W. Entonces podríamos tomar el calor Q_1 del depósito a T_1 y con la máquina B podríamos efectuar el trabajo W' y entregar algún calor al depósito a T_2; no nos importa cuánto. Hecho esto, podríamos ahorrar algo del trabajo W', que se supone mayor que W, podríamos usar una parte W de él y guardar el resto $W' - W$ para trabajo útil. Con el trabajo W podríamos hacer que la máquina A funcionase al revés *ya que es una máquina reversible*. Absorberá algún calor del depósito a T_2 y devolverá Q_1 al depósito a T_1. Después de este doble ciclo, el resultado neto sería que hemos puesto todas las cosas como las teníamos al principio y que hemos realizado algún trabajo adicional, $W' - W$, y todo lo que habríamos hecho sería ¡extraer energía del depósito a T_2! Tuvimos cuidado de devolver el calor Q_1 al depósito a T_1. Por lo tanto, dicho depósito puede ser pequeño y estar "dentro" de nuestra máquina combinada $A + B$, cuyo efecto neto es por lo demás extraer un calor neto $W' - W$ del depósito a T_2 y convertido en

trabajo. Pero obtener trabajo útil de un depósito a una sola temperatura *sin ningún otro cambio* es imposible de acuerdo al postulado de Carnot; no se puede hacer. Por consiguiente, ninguna máquina que absorbe una cantidad determinada de calor a una temperatura mayor T_1 y la entrega a la temperatura T_2 puede efectuar más trabajo que una máquina reversible trabajando bajo las mismas condiciones de temperatura.

Supongan ahora que la máquina B también es reversible. Entonces, naturalmente, no sólo no debe ser W' mayor que W, sino que ahora podemos invertir el razonamiento y demostrar que W no puede ser mayor que W'. Por lo tanto, si las dos máquinas son reversibles, ambas deben realizar la misma cantidad de trabajo y así llegamos a la brillante conclusión de Carnot: si una máquina es reversible, no importa cómo está diseñada, porque la cantidad de trabajo que se obtendrá si la máquina absorbe una determinada cantidad de calor a la temperatura T_1 y entrega calor a alguna otra temperatura T_2 *no depende en nada del diseño de la máquina.* Es una propiedad del mundo y no de una máquina en particular.

Si pudiésemos encontrar la ley que determina cuánto trabajo se obtiene cuando se absorbe el calor Q_1 de T_1 y se entrega calor a T_2, esta cantidad sería algo universal, independiente de la sustancia. Naturalmente que si conociésemos las propiedades de una sustancia particular, podríamos calcularlo y decir entonces que todas las demás sustancias dan la misma cantidad de trabajo en una máquina reversible. Ésta es la idea clave, la pista mediante la cual podríamos encontrar, por ejemplo, la relación entre cuánto se contrae una banda elástica cuando la calentamos y cuánto se enfría cuando la dejamos contraer. Imaginen que colocamos dicha banda en una máquina reversible y hacemos que complete un ciclo reversible. El resultado neto, cantidad total de trabajo efectuado, es esa función universal, esa gran función que es independiente de la sustancia. Vemos así que las propiedades de una sustancia deben estar limitadas de un cierto modo; no se puede hacer todo lo que uno quiere, o si no se podría inventar una sustancia que se pudiese usar para producir más que el máximo trabajo permitido cuando se llevase en un ciclo reversible. Este principio, esta limitación, es la única regla real que sale de la termodinámica.

44-4 Eficiencia de una máquina ideal

Ahora trataremos de encontrar la ley que determina el trabajo W en función de Q_1, T_1 y T_2. Es obvio que W es proporcional a Q_1, ya que si consideramos dos máquinas reversibles en paralelo, ambas trabajando juntas y ambas máquinas dobles, la combinación es también una máquina reversible. Si cada una absorbe el calor Q_1, las dos juntas absorben $2Q_1$ y el trabajo efectuado es $2W$, etc. Por lo que es razonable que W sea proporcional a Q_1.

El siguiente paso importante es encontrar esta ley universal. Podemos y queremos hacerlo estudiando una máquina reversible con una sustancia particular cuyas leyes conocemos, con un gas perfecto. También es posible obtener la regla mediante un razonamiento puramente lógico, sin usar de ningún modo una sustancia particular. Ésta es una de las piezas maestras del razonamiento físico y estamos deseosos de mostrársela, por lo que la discutiremos dentro de un momento para aquellos que deseen verla. Pero antes, usaremos el método mucho menos abstracto y simple del cálculo directo para un gas perfecto.

Necesitamos solamente obtener fórmulas para Q_1 y Q_2 (porque W es justamente $(Q_1 - Q_2)$, calores intercambiados con los depósitos durante la expansión o contracción isotérmicas. Por ejemplo, ¿cuánto calor Q_1 se ha tomado del depósito a temperatura T_1 durante la expansión isotérmica [marcada (1) en la figura 44-6] desde el punto a, a presión p_a, volumen V_a y temperatura T_1, hasta el punto b a presión p_b, volumen V_b y la misma temperatura T_1? Para un gas perfecto cada molécula tiene una energía que depende solamente de la temperatura, y puesto que la temperatura y el número de moléculas son iguales en a y b, la energía interna es la misma. *No hay cambio en* U, todo el trabajo efectuado por el gas,

$$W = \int_a^b p\,dV,$$

durante la expansión es energía Q_1 tomada del depósito. Durante la expansión, $pV = NkT_1$, o sea

$$p = \frac{NkT_1}{V}$$

o

$$Q_1 = \int_a^b p\,dV = \int_a^b NkT_1\,\frac{dV}{V} \tag{44.4}$$

o

$$Q_1 = NkT_1 \ln \frac{V_b}{V_a}$$

es el calor tomado del depósito a T_1. Del mismo modo, para la compresión a T_2 [curva (3) de la figura 44-6] el calor entregado al depósito a T_2 es

$$Q_2 = NkT_2 \ln \frac{V_c}{V_d}. \tag{44.5}$$

Para completar nuestro análisis necesitamos solamente encontrar una relación entre V_c/V_d y V_b/V_a. Esto lo hacemos observando que (2) es una expansión adiabáticas desde b hasta c y durante ella pV^γ es una constante. Como $pV = NkT$, podemos escribir esto en la forma $(pV)V^{\gamma-1}$ = constante o, en función de T y V, en la forma $TV^{\gamma-1}$ = constante, es decir

$$T_1V_b^{\gamma-1} = T_2V_c^{\gamma-1}. \tag{44.6}$$

Análogamente a (4), la compresión desde d hasta a es también adiabática, así

$$T_1V_a^{\gamma-1} = T_2V_d^{\gamma-1}. \tag{44.6a}$$

Si dividimos esta ecuación entre la anterior encontramos que V_b/V_a, debe ser igual a V_c/V_d, por lo que los logaritmos de (44.4) y (44.5) son iguales y resulta:

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2}. \tag{44.7}$$

Ésta es la relación que estábamos buscando. Aunque se ha probado para una máquina de gas perfecto, sabemos que debe ser válida *para cualquier máquina reversible*.

Ahora veremos cómo esta ley universal se pudo también haber obtenido mediante un razonamiento lógico, sin conocer las propiedades de una sustancia particular, como sigue. Supongan que tenemos tres máquinas y tres temperaturas, digamos T_1, T_2 y T_3. Digamos que una máquina absorbe el calor Q_1 de la temperatura T_1 y que efectúa una cierta cantidad de trabajo W_{13} y que entrega el calor Q_3 a la temperatura T_3 (figura 44-8). Supongamos que otra máquina va al revés entre T_2 y T_3, y que es de un tamaño tal que absorberá el mismo calor Q_3 y entregará Q_2. Tenemos que darle una cierta cantidad de trabajo W_{32} —que será negativo porque la máquina va al revés—. Cuando la primera máquina completa un ciclo, absorbe el calor Q_1 y entrega Q_3 a la temperatura T_3, pero la segunda máquina toma el mismo calor Q_3 del depósito a temperatura T_3 y lo entrega al depósito a temperatura T_2. Por lo tanto, el resultado neto de las dos máquinas en tándem es tomar el calor Q_1 de T_1 y entregar Q_2 a T_2. Las dos son, pues, equivalentes a una tercera que absorbe Q_1 a T_1, efectúa el trabajo W_{12} y entrega Q_2 a T_2, ya que $W_{12} = W_{13} - W_{32}$, como se puede demostrar inmediatamente a partir de la primera ley, como sigue:

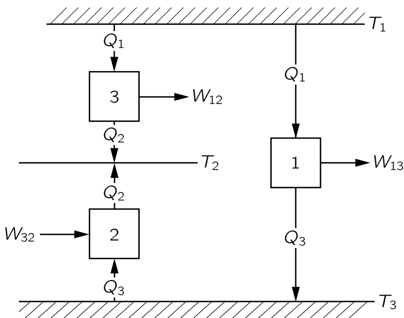


Figura 44-8. Las máquinas 1 y 2 juntas son equivalentes a la máquina 3.

$$W_{13} - W_{32} = (Q_1 - Q_3) - (Q_2 - Q_3) = Q_1 - Q_2 = W_{12}. \tag{44.8}$$

Ahora podemos obtener las leyes que relacionan las eficiencias de las máquinas, porque claramente debe haber alguna clase de relación entre las eficiencias de las máquinas que funcionan entre las temperaturas T_1 y T_3, y entre T_2 y T_3, y entre T_1 y T_2.

Podemos aclarar mucho el razonamiento del siguiente modo: acabamos de ver que siempre podemos relacionar el calor absorbido a T_1 con el entregado a T_2, encontrando el calor entregado a cualquier otra temperatura T_3. Por consiguiente, podemos obtener todas las propiedades de las máquinas si introducimos una temperatura patrón y analizándolo todo con ella. En otras palabras, si conociésemos la eficiencia de una máquina que funciona entre una cierta temperatura T y una cierta temperatura patrón arbitraria, podríamos calcular la eficiencia para cualquier otra diferencia de temperatura. Como suponemos que estamos usando solamente máquinas reversibles, podemos trabajar bajando desde la temperatura inicial a la temperatura patrón y subir de nuevo a la temperatura final. Definiremos la temperatura patrón arbitrariamente como de *un grado*. También adoptaremos un símbolo especial para el calor que es entregado a esta temperatura patrón: lo llamaremos Q_S. En otras palabras, cuando una máquina reversible absorbe el calor Q a la temperatura T, entregará, a la temperatura unitaria, un calor Q_S. *Si una máquina que absorbe Q_1 a T_1 entrega el calor Q_S a un grado, y si una máquina que absorbe Q_2 a la temperatura T_2 entrega también la misma cantidad de calor Q_S a un grado, se sigue que una máquina que absorbe el calor Q_1 a la temperatura T_1 entregará el calor Q_2 si funciona entre T_1 y T_2*, como ya habíamos demostrado considerando máquinas que funcionan entre tres temperaturas. Por lo que todo lo que tenemos realmente que hacer es encontrar cuánto calor Q_1 necesitamos dar a la temperatura T_1, para entregar una cierta cantidad de calor Q_S a la temperatura unitaria. Si descubrimos esto, lo tenemos todo. El calor Q es, por supuesto, una función de la temperatura T. Es fácil ver que el calor debe aumentar cuando lo hace la temperatura, porque sabemos que se necesita trabajo para que una máquina funcione al revés y entregue calor a una temperatura más alta. Es también fácil ver que el calor Q_1 debe ser proporcional a Q_S. Así, la gran ley es algo como esto: para una cantidad dada de calor Q_S entregada a un grado por una máquina que funciona a la temperatura de T grados, el calor Q absorbido debe ser esa cantidad Q_S por una función creciente de la temperatura,

$$Q = Q_S f(T). \tag{44.9}$$

44-5 Temperatura termodinámica

A estas alturas no vamos a tratar de encontrar la fórmula para esa función creciente de la temperatura en términos de nuestra familiar escala de temperatura de mercurio, sino que *definiremos la temperatura mediante una nueva escala*. En un tiempo se definía “la temperatura” arbitrariamente dividiendo la dilatación del agua en grados iguales de un cierto tamaño. Pero cuando se mide la temperatura con un termómetro de mercurio se encuentra que ya los grados no son iguales. Pero *ahora podemos dar una definición de temperatura que es independiente de cualquier sustancia particular*. Podemos usar aquella función $f(T)$ que no depende de lo que usemos, ya que la eficiencia de estas máquinas reversibles es independiente de las sustancias con que trabajan. Como la función que encontramos aumenta con la temperatura, definiremos la *función misma* como temperatura, medida en unidades de la temperatura patrón de un grado, en la forma siguiente:

$$Q = ST, \tag{44.10}$$

donde

$$Q_S = S \cdot 1^\circ. \tag{44.11}$$

Esto significa que podemos decir lo caliente que está un objeto encontrando cuánto calor es absorbido por una máquina reversible que trabaja entre la temperatura, el objeto y la temperatura unitaria (figura 44-9). Si se toma de una caldera siete veces más calor del que es entregado a un condensador de un grado, la temperatura de la caldera se dirá que es de siete grados, y así sucesivamente. Por consiguiente, midiendo cuánto calor es absorbido a diferentes temperaturas,

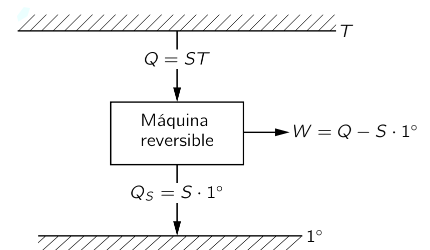


Figura 44-9. Temperatura termodinámica absoluta.

determinamos la temperatura. La temperatura definida de este modo se denomina *temperatura termodinámica absoluta*, y es independiente de la sustancia. Desde ahora en adelante usaremos exclusivamente esta definición.*

Ahora nos damos cuenta de que cuando tenemos dos máquinas, una trabajando entre T_1 y un grado, la otra entre T_2 y un grado, y entregan el mismo calor a la temperatura unitaria, los calores absorbidos deben estar relacionados mediante

$$\frac{Q_1}{T_1} = S = \frac{Q_2}{T_2}. \tag{44.12}$$

Pero esto significa que si tenemos una sola máquina que funciona entre T_1 y T_2, entonces el resultado del análisis total, el gran final, es que Q_1 es a T_1 como Q_2 es a T_2 si la máquina absorbe Q_1 a la temperatura T_1 y entrega el calor Q_2 a la temperatura T_2. Siempre que la máquina sea reversible, se debe seguir esta relación entre los calores. Esto es todo lo que hay, éste es el centro del universo termodinámico.

Si todo esto es lo que hay en termodinámica, ¿por qué se la considera una materia tan difícil? Al resolver un problema que envuelve una masa dada de una sustancia, el estado de la sustancia en cualquier momento se puede describir diciendo cuál es su temperatura y cuál es su volumen. Si conocemos la temperatura y el volumen de una sustancia, y que la presión es una cierta función de la temperatura y del volumen, conocemos entonces la energía interna. Alguien podría decir: "Yo no quiero hacerlo así. Dígame la temperatura y la presión y yo le diré el volumen. Yo puedo considerar el volumen como una función de la temperatura y presión, y la energía interna como una función de la temperatura y presión, etc." Ésta es la razón por la cual la termodinámica es difícil, porque cada uno usa un enfoque diferente. Si solamente pudiésemos reunirnos por una sola vez y decidir cuáles van a ser nuestras variables y aceptarlo así siempre, sería bastante fácil.

Comencemos ahora a hacer deducciones. Así como $F = ma$ es el centro del universo en mecánica, y sigue y sigue y sigue después de esto, el principio que acabamos de encontrar lo es todo para la termodinámica. Pero, ¿se pueden sacar conclusiones de él?

Comenzamos. Para obtener nuestra primera conclusión, combinamos ambas leyes, la ley de la conservación de la energía y esta ley que relaciona los calores Q_1 y Q_2 y podemos obtener fácilmente *la eficiencia de una máquina reversible*. Según la primera ley, tenemos que $W = Q_1 - Q_2$. Según nuestro nuevo principio,

$$Q_2 = \frac{T_2}{T_1}\,Q_1,$$

por lo que el trabajo resulta ser

$$W = Q_1\left(1 - \frac{T_2}{T_1}\right) = Q_1\,\frac{T_1 - T_2}{T_1}, \tag{44.13}$$

que nos dice la eficiencia de la máquina —cuánto trabajo obtenemos con tanto calor—. La eficiencia de una máquina es proporcional a la diferencia de temperaturas entre las que trabaja, dividido entre la temperatura mayor:

$$\text{Eficiencia} = \frac{W}{Q_1} = \frac{T_1 - T_2}{T_1}. \tag{44.14}$$

La eficiencia no puede ser mayor que la unidad y la temperatura absoluta no puede ser menor que cero, cero absoluto. Así, pues, como T_2 debe ser positiva, la eficiencia es siempre menor que la unidad. Ésta es nuestra primera conclusión.

* Hemos definido previamente nuestra escala de temperatura de un modo diferente, estableciendo que la energía cinética media de una molécula en un gas perfecto es proporcional a la temperatura, o sea que la ley del gas perfecto dice que pV es proporcional a T. ¿Es equivalente esta nueva definición? Sí, puesto que el resultado final (44.7) derivado de la ley del gas es igual que el derivado aquí. Volveremos a discutir este punto en el siguiente capítulo.

44-6 Entropía

Las ecuaciones (44.7) y (44.12) pueden ser interpretadas de una manera especial. Trabajando siempre con máquinas reversibles, un calor Q_1 a la temperatura T_1 es "equivalente" a Q_2 a T_2 si $Q_1/T_1 = Q_2/T_2$, en el sentido de que cuando se absorbe uno se entrega el otro. Esto sugiere que si le damos algún nombre a Q/T, podemos decir: en un proceso reversible se absorbe tanto Q/T como se libera, no hay ganancia o pérdida de Q/T. Este Q/T se llama *entropía* y entonces decimos que "no hay variación neta de entropía en un ciclo reversible". Si T es 1°, la entropía es $Q_S/1°$ o, usando el símbolo que le dimos, $Q_S/1° = S$. Comúnmente S es la letra usada para entropía y es numéricamente igual al calor (que hemos llamado Q_S) entregado a un depósito de 1° (la entropía no es un calor, es un calor dividido entre una temperatura, por lo que se mide en *joules por grado*).

Ahora bien, es interesante que además de la presión y la energía interna, que son funciones de la temperatura y del volumen, hayamos encontrado otra cantidad que es una función del estado, esto es, la entropía de la sustancia. Tratemos de explicar cómo la calculamos y qué queremos decir cuando la llamamos una "función del estado". Consideren el sistema en dos estados diferentes, muy parecido a lo que teníamos en el experimento donde realizamos las expansiones adiabáticas e isotérmicas. (Entre paréntesis, no hay necesidad de que una máquina térmica tenga solamente dos depósitos, podría tener tres o cuatro temperaturas diferentes a las que toma y entrega calor, etc.) Podemos movernos sobre un diagrama pV en todo momento e ir de un estado a otro. En otras palabras, podríamos decir que el gas está en un cierto estado a y luego pasa a otro b, y nosotros exigiremos que esta transición, de a a b, sea reversible. Supongan ahora que a todo lo largo del camino desde a hasta b tenemos pequeños depósitos a temperaturas diferentes, de modo que el calor dQ sacado de la sustancia en cada uno de los pequeños pasos se entrega a cada depósito a la temperatura correspondiente al punto del camino. Conectamos entonces estos depósitos, mediante máquinas térmicas reversibles, a un solo depósito a la temperatura unitaria. Cuando hayamos terminado de llevar la sustancia desde a hasta b, llevaremos todos los depósitos a su estado inicial. Cualquier cantidad de calor dQ que se ha absorbido de la sustancia a la temperatura T ha sido convertida ahora por una máquina reversible y se ha entregado una cierta cantidad de entropía dS a la temperatura unitaria:

$$dS = dQ/T. \tag{44.15}$$

Calculemos la cantidad total de entropía que ha sido entregada. La diferencia de entropía, o entropía necesaria para ir de a a b mediante esta transformación reversible particular, es la entropía total, el total de entropía que se ha tomado de los pequeños depósitos y que se ha entregado a la temperatura unitaria:

$$S_b - S_a = \int_a^b \frac{dQ}{T}. \tag{44.16}$$

Hay una pregunta: ¿depende la diferencia de entropía del camino seguido? Existe más de un camino para ir de a a b. Recuerden que en el ciclo de Carnot podemos ir de a a c (véase la figura 44-6) expandiendo primero isotérmicamente y luego adiabáticamente, o podemos expandir primero adiabáticamente y luego isotérmicamente. La pregunta es si la variación de entropía que tiene lugar cuando vamos de a a b, en la figura 44-10, es la misma según sigamos un camino u otro. *Debe ser igual*, puesto que si hiciésemos el ciclo completo, yendo hacia adelante por un camino y hacia atrás por el otro, tendríamos una máquina reversible y no habría pérdida de calor en el depósito a temperatura unitaria. En un ciclo totalmente reversible, no se debe tomar ningún calor del depósito a temperatura unitaria, por lo que la entropía necesaria para ir de a a b es igual siguiendo diferentes caminos. *Es independiente del camino* y depende solamente de los puntos extremos. Podemos decir, por consiguiente, que hay una cierta función, que llamamos entropía de la sustancia, que depende solamente del estado, es decir, sólo del volumen y de la temperatura.

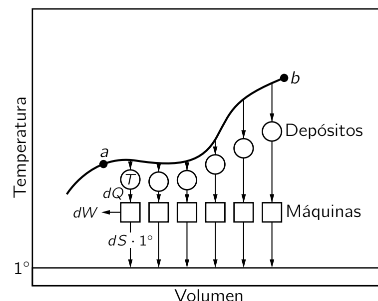


Figura 44-10. Variación de entropía durante una transformación reversible.

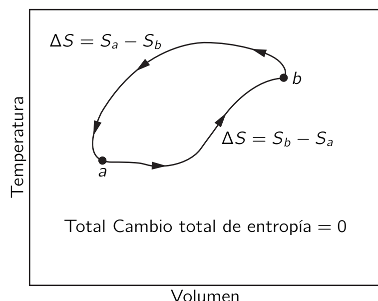


Figura 44-11. Variación de entropía en un ciclo completamente reversible.

Podemos encontrar una función $S(V, T)$ que tiene la propiedad de que si calculamos la variación de entropía en términos del calor dejado a la temperatura unitaria, cuando la sustancia se mueve según cualquier camino reversible, entonces

$$\Delta S = \int \frac{dQ}{T}, \tag{44.17}$$

donde dQ es el calor sacado de la sustancia a la temperatura T. Esta variación total de entropía es la diferencia entre la entropía calculada en los puntos inicial y final:

$$\Delta S = S(V_b, T_b) - S(V_a, T_a) = \int_a^b \frac{dQ}{T}. \tag{44.18}$$

Esta expresión no define completamente la entropía, sino más bien solamente la *diferencia* de entropía entre dos estados diferentes. Únicamente si podemos evaluar la entropía para un estado especial, podemos en realidad definir S en forma absoluta.

Durante mucho tiempo se creyó que la entropía absoluta no significaba nada —que sólo se podrían definir diferencias—, pero por fin Nernst propuso lo que llamó *teorema del calor*, que también se llama tercera ley de la termodinámica. Es muy sencillo. Diremos en qué consiste, pero no explicaremos por qué es válido. El postulado de Nernst afirma simplemente que la entropía de cualquier objeto es cero en el cero absoluto. Sabemos de un caso de T y V, cuando $T = 0$, donde S es cero; y así podemos obtener la entropía en cualquier otro punto.

Para ilustrar estas ideas, calculemos la entropía de un gas perfecto. En una expansión isotérmica (y, por lo tanto, reversible), $\int dQ/T$ es Q/T, puesto que T es constante. Por consiguiente, según (44.4), la variación de entropía es

$$S(V_a, T) - S(V_b, T) = Nk \ln \frac{V_a}{V_b},$$

y por lo tanto $S(V, T) = Nk \ln V$ más alguna función de T únicamente. ¿Cómo depende S de T? Sabemos que para una expansión adiabática reversible, *no se intercambia calor*. Luego, la entropía no cambia aunque V lo haga, siempre que T también cambie, de tal forma que $TV^{\gamma-1}$ = constante. ¿Pueden ver que esto implica que:

$$S(V, T) = Nk\left[\ln V + \frac{1}{\gamma - 1}\ln T\right] + a,$$

donde a es una constante independiente tanto *de V como de T*? (a se llama constante química. Depende del gas en cuestión y se puede determinar experimentalmente a partir del teorema de Nernst midiendo el calor liberado al enfriar y condensar el gas hasta que ha pasado a sólido —o en el caso del helio, a líquido— a 0°, integrando $\int dQ/T$. También se puede determinar teóricamente mediante la constante de Planck y la mecánica cuántica, pero no lo estudiaremos en este curso).

Ahora señalaremos algunas propiedades de la entropía de las cosas. Recordemos, en primer lugar, que si vamos según un ciclo reversible de a a b, la entropía de la sustancia cambiará en $S_b - S_a$. Y también recordamos que a medida que recorremos el camino, la entropía —calor entregado a la temperatura unitaria— aumenta de acuerdo con la regla $dS = dQ/T$, donde dQ es el calor que quitamos de la sustancia cuando su temperatura es T.

Ya sabemos que si tenemos un *ciclo* reversible, la entropía total de todo no varía, ya que el calor Q_1 absorbido a T_1 y el calor Q_2 entregado a T_2 corresponden a variaciones iguales y opuestas de entropía, por lo que la variación neta de entropía es cero. Por tanto, en un ciclo reversible no hay cambio en la entropía de nada, incluyendo los depósitos. Esta regla se puede parecer a la conservación de la energía, pero no lo es, se aplica solamente a ciclos reversibles. Si incluimos ciclos irreversibles no hay ley de conservación de la entropía.

Daremos dos ejemplos. En primer lugar, supongan que hacemos trabajo irreversible sobre un objeto mediante fricción, generando un calor Q en algún

objeto a temperatura T. La entropía aumenta en Q/T. El calor Q es igual al trabajo y así cuando efectuamos una cierta cantidad de trabajo por fricción contra un objeto cuya temperatura es T, la entropía de todo el mundo aumenta en W/T.

Otro ejemplo de irreversibilidad es éste: si juntamos dos objetos que están a temperaturas diferentes, T_1 y T_2 digamos, una cierta cantidad de calor fluirá del uno al otro espontáneamente. Supongan, por ejemplo, que colocamos una piedra caliente en agua fría. Entonces, cuando una cierta cantidad de calor ΔQ ha pasado de T_1 a T_2, ¿cuánto cambia la entropía de la piedra caliente? Disminuye en $\Delta Q/T_1$. ¿Cuánto cambia la entropía del agua? Aumenta en $\Delta Q/T_2$. El calor fluirá, naturalmente, sólo desde la temperatura más alta T_1 a la más baja T_2, por lo que ΔQ es positiva si T_1 es mayor que T_2. Así, pues, la variación de entropía de todo el universo es positiva y es la diferencia entre las dos fracciones:

$$\Delta S = \frac{\Delta Q}{T_2} - \frac{\Delta Q}{T_1}. \tag{44.19}$$

Vale entonces la siguiente proposición: la entropía de todo el universo aumenta en cualquier proceso que es irreversible. Solamente en procesos reversibles la entropía sí permanece constante. Como ningún proceso es absolutamente reversible, siempre hay por lo menos una pequeña ganancia en la entropía; un proceso reversible es una idealización en la que hemos hecho mínima la ganancia de entropía.

Desafortunadamente, no vamos a ir muy lejos dentro del campo de la termodinámica. Nuestro propósito es solamente ilustrar las ideas principales involucradas, así como las razones de por qué es posible hacer tales argumentos, pero no usaremos mucho la termodinámica en este curso. Los ingenieros y particularmente los químicos usan la termodinámica con mucha frecuencia. Así que debemos aprender nuestra termodinámica práctica en química o en ingeniería. Como no vale la pena repetir las cosas, solamente discutiremos algo del origen de la teoría en lugar de entrar en detalles de aplicaciones particulares.

Con frecuencia las dos leyes de la termodinámica se enuncian así:

Primera ley: la energía del universo es siempre constante.

Segunda ley: la entropía del universo siempre está aumentando.

No es un buen enunciado el de la segunda ley; no dice, por ejemplo, que en un ciclo reversible la entropía permanece constante, ni tampoco dice exactamente qué es la entropía. Es sólo un modo ingenioso de recordar las dos leyes, pero no nos dice en realidad exactamente de qué tratan. Hemos resumido las leyes discutidas en este capítulo en la tabla 44-1. En el próximo capítulo aplicaremos estas leyes para descubrir la relación entre el calor generado en la expansión de una banda elástica y la tensión adicional cuando se la calienta.

Tabla 44-1

Resumen de las leyes de la termodinámica

Primera ley:

Calor dado a un sistema + Trabajo efectuado sobre el sistema = Aumento en la energía interna del sistema:

$$dQ + dW = dU.$$

Segunda ley:

Es imposible un proceso cuyo *único* resultado final sea tomar calor de un depósito y convertirlo en trabajo.

Ninguna máquina térmica que toma calor Q_1 de T_1 y que entrega calor Q_2 a T_2 puede efectuar más trabajo que una máquina reversible, para la cual

$$W = Q_1 - Q_2 = Q_1\left(\frac{T_1 - T_2}{T_1}\right).$$

La entropía de un sistema se define así:

(a) Si se añade reversiblemente el calor ΔQ a un sistema a temperatura T, el aumento de entropía del sistema es $\Delta S = \Delta Q/T$.

(b) En $T = 0$, $S = 0$ (*tercera ley*).

En un *cambio reversible*, la entropía total de todas las partes del sistema (incluyendo depósitos) no varía.

En un *cambio irreversible*, la entropía total del sistema siempre aumenta.

45

Ejemplos de termodinámica

45-1 Energía interna

La termodinámica es bastante difícil y compleja cuando llegamos a aplicarla y no nos compete en este curso profundizar mucho en las aplicaciones. El tema es de muchísima importancia, naturalmente, para ingenieros y químicos, y los que estén interesados en el mismo pueden enterarse de las aplicaciones en fisicoquímica y en termodinámica para ingenieros. Hay también buenos libros de referencia, como *Calor y termodinámica* de Zemansky, donde se puede profundizar el tema. En la *Encyclopædia Britannica*, decimocuarta edición, se pueden encontrar artículos excelentes sobre termodinámica y termoquímica, y en el artículo sobre química, las secciones de fisicoquímica, vaporización, licuefacción de gases y otros.

La termodinámica es complicada porque hay muchas maneras diferentes de describir la misma cosa. Si deseamos describir el comportamiento de un gas, podemos decir que la presión depende de la temperatura y del volumen o que el volumen depende de la temperatura y la presión. Si consideramos la energía interna U, podríamos decir que depende de la temperatura y el volumen, si ésas son las variables que hemos elegido, pero también podríamos decir que depende de la temperatura y la presión, o la presión y el volumen, y así sucesivamente. En el último capítulo discutimos otra función de la temperatura y el volumen llamada entropía S, y desde luego podemos construir tantas funciones de esas variables como queramos: $U - TS$ es una función de la temperatura y el volumen. Tenemos, pues, un gran número de cantidades diferentes que puedan ser funciones de muchas combinaciones distintas de variables.

Para mantenernos dentro de ciertos límites de simplicidad en este capítulo, decidiremos desde el principio usar *temperatura* y *volumen* como variables independientes. Los químicos usan temperatura y presión porque son más fáciles de medir y de controlar en los experimentos químicos; nosotros, en cambio, usaremos temperatura y volumen en todo este capítulo, excepto en un punto donde veremos cómo se hace la transformación al sistema de variables usado por los químicos.

En primer lugar, consideraremos entonces únicamente un sistema de variables independientes, temperatura y volumen. En segundo lugar, sólo discutiremos dos funciones dependientes, la energía interna y la presión. Todas las otras funciones se pueden obtener a partir de éstas, por lo que no es necesario discutirlas. Aun con estas limitaciones, la termodinámica es un tema bastante difícil ¡pero no del todo imposible!

Recordemos primero un poco de matemática. Si una cantidad es función de dos variables, el concepto de derivada de una cantidad requiere un examen más cuidadoso que en el caso en que hay una sola variable. ¿Qué entendemos por derivada de la presión respecto a la temperatura? El cambio de presión que acompaña a una modificación de temperatura depende en parte, por supuesto, de lo que le ocurre al *volumen* mientras T cambia. Debemos especificar la variación de V para que el concepto de derivada respecto a T tenga un significado preciso. Podría interesarnos, por ejemplo, la rapidez de variación de P con T si se mantiene V constante. Este cociente es simplemente la derivada ordinaria que escribimos comúnmente como dP/dT. Habitualmente, se usa un símbolo especial, $\partial P/\partial T$, para recordar que P depende de otra variable V además de T, y que esa otra variable se mantiene constante. No sólo usaremos el símbolo ∂ para llamar la atención sobre el hecho de que la otra variable se mantiene constante, sino que también escribiremos como subíndice la variable que se mantiene constante, $(\partial P/\partial t)_V$. Como tenemos sólo dos variables independientes, esta notación es

redundante, pero nos servirá para no perder el camino de la jungla termodinámica de derivadas parciales.

Supongamos que la función $f(x, y)$ depende de las dos variables independientes x y y. Indicamos con $(\partial f/\partial x)_y$ la derivada ordinaria obtenida del modo habitual, tratando y como constante:

$$\left(\frac{\partial f}{\partial x}\right)_y = \underset{\Delta x \to 0}{\text{límite}} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}.$$

Análogamente definimos

$$\left(\frac{\partial f}{\partial y}\right)_x = \underset{\Delta y \to 0}{\text{límite}} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}.$$

Por ejemplo, si $f(x, y) = x^2 + yx$, se tiene $(\partial f/\partial x)_y = 2x + y$, y $(\partial f/\partial y)_x = x$. Podemos extender este concepto a derivadas de orden superior: $\partial^2 f/\partial y^2$ o $\partial^2 f/\partial y \partial x$. El último símbolo indica que primero derivamos f respecto a x tratando y como constante y luego derivamos el resultado respecto a y tratando x como constante. El orden de derivación no importa: $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$.

Será necesario calcular el incremento Δf de $f(x, y)$ cuando x pasa a $x + \Delta x$ *al mismo tiempo* que y pasa a $y + \Delta y$. Supondremos en lo que sigue que Δx y Δy son infinitesimales:

$$\begin{aligned} \Delta f &= f(x + \Delta x, y + \Delta y) - f(x, y) \\ &= \underbrace{f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y)} + \underbrace{f(x, y + \Delta y) - f(x, y)} \\ &= \qquad \Delta x \left(\frac{\partial f}{\partial x}\right)_y \qquad + \qquad \Delta y \left(\frac{\partial f}{\partial y}\right)_x \end{aligned} \tag{45.1}$$

La última ecuación es la relación fundamental que expresa Δf en términos de Δx y Δy.

Como ejemplo del uso de esta relación, calculemos el incremento de la energía interna $U(T, V)$ cuando la temperatura varía de T a $T + \Delta T$ y el volumen varía de V a $V + \Delta V$. Empleando la ecuación (45.1) escribimos:

$$\Delta U = \Delta T \left(\frac{\partial U}{\partial T}\right)_V + \Delta V \left(\frac{\partial U}{\partial V}\right)_T. \tag{45.2}$$

En el capítulo anterior encontramos otra expresión para el incremento ΔU de la energía interna cuando se agrega al gas una cantidad de calor ΔQ:

$$\Delta U = \Delta Q - P\,\Delta V. \tag{45.3}$$

Comparando las ecuaciones (45.2) y (45.3) se podría creer a primera vista que $P = -(\partial U/\partial V)_T$, pero esto no es correcto. Para obtener la relación correcta, supongamos primero que agregamos una cantidad de calor ΔQ al gas manteniendo el volumen constante, de modo que $\Delta V = 0$. Si $\Delta V = 0$, la ecuación (45.3) nos dice que $\Delta U = \Delta Q$ y la ecuación (45.2) nos dice que $\Delta U = (\partial U/\partial T)_V \Delta T$ de donde $(\partial U/\partial T)_V = \Delta Q/\Delta T$. El cociente $\Delta Q/\Delta T$, cantidad de calor que se debe entregar a una sustancia para cambiar su temperatura en un grado manteniendo el volumen constante, se llama *calor específicoa volumen constante* y se indica con el símbolo C_V. Hemos demostrado, pues, que:

$$\left(\frac{\partial U}{\partial T}\right)_V = C_V. \tag{45.4}$$

Ahora agreguemos de nuevo al gas una cantidad de calor ΔQ, pero esta vez manteniendo T constante y dejando que el volumen cambie en ΔV. En este caso el análisis es más complejo, pero podemos calcular ΔU con el razonamiento de Carnot, haciendo uso del ciclo de Carnot que introdujimos en el capítulo precedente.

La figura 45-1 muestra el diagrama presión-volumen para el ciclo de Carnot. Como ya hemos visto, el trabajo total realizado por el gas en un ciclo reversible

es $\Delta Q(\Delta T/T)$, donde ΔQ es la cantidad de energía térmica suministrada al gas a medida que se expande isotérmicamente a la temperatura T del volumen V al $V+\Delta V$, y $T-\Delta T$ es la temperatura final que alcanza el gas cuando se expande adiabáticamente en el segundo tramo del ciclo. Demostraremos ahora que este trabajo realizado también está dado por el área sombreada de la figura 45-1. En cualquier caso, el trabajo realizado por el gas es $\int P\,dV$ y es positivo cuando el gas se expande y negativo cuando el gas se comprime. Si hacemos el diagrama de P en función de V, la variación de P y V está representada por una curva cuyo valor de P correspondiente a un valor determinado de V. A medida que el volumen cambia de un valor a otro, el trabajo realizado por el gas, o sea la integral $\int P\,dV$, es el área que está debajo la curva que une los valores inicial y final de V. Aplicando esta idea al ciclo de Carnot, vemos que al recorrer el ciclo poniendo atención en el signo del trabajo realizado por el gas, el trabajo total realizado por el gas es justamente el área sombreada de la figura 45-1.

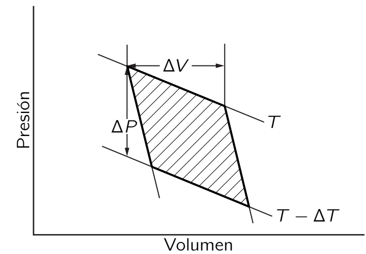


Figura 45-1. Diagrama de presión-volumen para un ciclo de Carnot. Las curvas indicadas con T y con $T-\Delta T$ son isotermas, las curvas más empinadas son adiabáticas. ΔV es la variación del volumen cuando se agrega el calor ΔQ al gas a temperatura constante T. ΔT es la variación de presión a volumen constante que resulta cuando la temperatura del gas varía de T a $T-\Delta T$.

Ahora calcularemos geométricamente el área sombreada. El ciclo usado en la figura 45-1 difiere del usado en el capítulo anterior en que ahora suponemos que ΔT y ΔQ son infinitesimales. Estamos trabajando entre adiabáticas e isotermas muy cercanas, y la figura formada por las líneas gruesas de la figura 45-1 tenderá a un paralelogramo cuando los incrementos ΔT y ΔQ tiendan a cero. El área de este paralelogramo es $\Delta V\Delta P$, donde ΔV es la variación de volumen que resulta de entregar la energía ΔQ al gas a temperatura constante, y ΔP es la variación de presión cuando la temperatura varía en ΔT a volumen constante. Se puede demostrar fácilmente que el área sombreada de la figura 45-1 está dada por $\Delta V\Delta P$ observando que es equivalente al área encerrada por las líneas de trazos de la figura 45-2, que a su vez difiere del rectángulo formado por ΔP y ΔV únicamente por la suma y la resta de las áreas triangulares iguales de la figura 45-2.

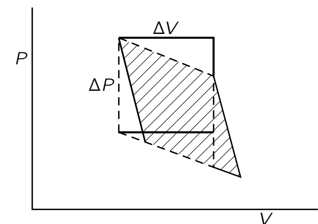


Figura 45-2. Área som. = área encerrada por las líneas de trazos = área del rectángulo = $\Delta P\Delta V$.

Resumamos los resultados obtenidos hasta ahora:

$$\left.\begin{aligned}
&\text{Trabajo realizado por el gas} = \text{área som.} = \Delta V\,\Delta P = \Delta Q\left(\frac{\Delta T}{T}\right)\\
\text{o}\quad&\\
&\frac{\Delta T}{T}\cdot(\text{calor necesario para cambiar } V \text{ en } \Delta V)_{\text{constante } T}\\
&\quad = \Delta V\cdot(\text{variación de } P \text{ cuando } T \text{ varía } \Delta T)_{\text{constante } V}\\
\text{o}\quad&\\
&\frac{1}{\Delta V}\cdot(\text{calor necesario para cambiar } V \text{ en } \Delta V)_T = T(\partial P/\partial T)_V.
\end{aligned}\right\}\tag{45.5}$$

La ecuación (45.5) expresa el resultado esencial del razonamiento de Carnot. Toda la termodinámica se puede deducir de la ecuación (45.5) y de la primera ley, expresada en la ecuación (45.3). La ecuación (45.5) es esencialmente la segunda ley, aunque Carnot la dedujo originariamente en una forma ligeramente diferente, ya que no utilizó nuestra definición de temperatura.

Ahora podemos proceder al cálculo de $(\partial U/\partial V)_T$. ¿En cuánto variaría la energía interna U si el volumen variase en ΔV? Primero: U varía porque se agrega calor, y segundo: U varía porque se realiza trabajo. El calor agregado es:

$$\Delta Q = T\left(\frac{\partial P}{\partial T}\right)_V\Delta V,$$

de acuerdo con la ecuación (45.5) y el trabajo realizado sobre la sustancia es $-P\Delta V$. En consecuencia, la variación ΔU de la energía interna tiene dos partes:

$$\Delta U = T\left(\frac{\partial P}{\partial T}\right)_V\Delta V - P\,\Delta V.\tag{45.6}$$

Dividiendo ambos miembros por ΔV, encontramos para la variación específica de U con V a T constante

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial P}{\partial T}\right)_V - P. \tag{45.7}$$

En nuestra presentación de la termodinámica, en la cual T y V son las únicas variables y P y U las únicas funciones, las ecuaciones (45.3) y (45.7) son las ecuaciones fundamentales de las que se pueden deducir todos los resultados.

45-2 Aplicaciones

Discutamos ahora el significado de la ecuación (45.7) y veamos por qué nos da la respuesta a las preguntas formuladas en el capítulo anterior. Consideramos el siguiente problema: en la teoría cinética es evidente que un aumento de temperatura induce un aumento de presión debido al bombardeo de los átomos sobre un pistón. Por la misma razón física, cuando hacemos retroceder el pistón se extrae calor del gas, y para mantener la temperatura constante hay que restituir calor al gas. El gas se enfría al expandirse y la presión aumenta al calentarse. Tiene que haber alguna conexión entre ambos fenómenos y la misma está dada explícitamente por la ecuación (45.7). Si mantenemos el volumen fijo y aumentamos la temperatura, la presión aumenta a razón de $(\partial P/\partial T)_V$. El hecho que sigue está relacionado con esto; si aumentamos el volumen, el gas se enfriará a no ser que agreguemos calor para mantener la temperatura constante y $(\partial U/\partial V)_T$ dice cuánto calor se necesita para conservar la temperatura. La ecuación (45.7) expresa la interrelación fundamental entre estos dos efectos. Esto es lo que prometimos encontrar cuando iniciamos el estudio de las leyes de la termodinámica. Sin conocer los mecanismos internos del gas y sabiendo únicamente que no puede haber movimiento perpetuo de segunda especie, ¡podemos deducir la relación entre la cantidad de calor necesaria para mantener constante la temperatura cuando el gas se expande, y la variación de presión cuando el gas se calienta a volumen constante!

Ahora que tenemos el resultado que queríamos en lo que respecta a un gas, consideremos el de la banda elástica. Cuando estiramos una banda elástica encontramos que su temperatura aumenta, y cuando la calentamos, se encoge. ¿Cuál es la ecuación que expresa, para una banda elástica, la misma relación que expresa la ecuación (45.3) para un gas? En el caso de una banda elástica, la situación es más o menos como sigue: cuando se agrega una cantidad de calor ΔQ, la energía interna cambia en ΔU y se realiza un trabajo. La única diferencia será que el trabajo realizado por la banda elástica es $-F\Delta L$ en vez de $P\Delta V$, donde F es la fuerza que se ejerce sobre la banda y L la longitud de la banda. La fuerza F es función de la temperatura y de la longitud de la banda. Remplazando $P\Delta V$ por $-F\Delta L$ en la ecuación (45.3), obtenemos:

$$\Delta U = \Delta Q + F\,\Delta L. \tag{45.8}$$

Comparando las ecuaciones (45.3) y (45.8) vemos que la ecuación de la banda elástica se obtiene por simple sustitución de una letra por otra. Más aún, si sustituimos V por L y P por $-F$, toda nuestra discusión del ciclo de Carnot sirve para la banda elástica. Por ejemplo, podemos deducir inmediatamente que el calor ΔQ necesario para cambiar la longitud en ΔL está dado por la análoga de la ecuación (45.5) $\Delta Q = -T(\partial F/\partial T)_L \Delta L$. Esta ecuación nos dice que si mantenemos fija la longitud de una banda elástica y la calentamos, podemos calcular cuánto aumentará la fuerza en función del calor necesario para mantener la temperatura constante cuando la banda se expande un poquito. Vemos, pues, que la misma ecuación sirve tanto para un gas como para una banda elástica. Verdaderamente, si se puede escribir $\Delta U = \Delta Q + A\Delta B$, donde A y B representan diversas cantidades, fuerza y longitud, presión y volumen, etc., se pueden aplicar los resultados obtenidos para un gas sustituyendo A y B por $-P$ y V. Por ejemplo, consideremos la diferencia de potencial eléctrico o "voltaje" E de una batería y la carga ΔZ que atraviesa la batería. Sabemos que el trabajo realizado en una celda eléctrica reversible, como un acumulador, es $E\Delta Z$. (Como

no incluimos un término $P\Delta V$ en el trabajo, requerimos que nuestra batería mantenga el volumen constante.) Veamos qué nos puede decir la termodinámica sobre el funcionamiento de una batería. Si remplazamos P por E y V por Z en la ecuación (45.6), obtenemos:

$$\frac{\Delta U}{\Delta Z} = T\left(\frac{\partial E}{\partial T}\right)_Z - E. \tag{45.9}$$

La ecuación (45.9) dice que la energía interna U cambia cuando una carga ΔZ atraviesa la celda. ¿Por qué $\Delta U/\Delta Z$ no es simplemente el voltaje E de la batería? (La respuesta es que una batería real se calienta cuando la atraviesa una carga. La energía interna de la batería cambia, primero porque la batería realizó un trabajo sobre el circuito externo, y segundo porque la batería se calienta.) Lo notable es que también se puede expresar la segunda parte en función de la manera en que el voltaje de la batería cambia con la temperatura. Entre paréntesis, cuando la carga atraviesa la batería hay reacciones químicas y la ecuación (45.9) sugiere un modo elegante de medir la cantidad de energía necesaria para producir una reacción química. ¡Todo lo que tenemos que hacer es construir una celda cuyo funcionamiento se base en esa reacción, medir el voltaje y medir cuánto cambia éste con relación a la temperatura cuando no extraemos cargas de la batería!

Ahora bien, hemos supuesto que se puede mantener constante el volumen de la batería; ya que hemos omitido el término $P\Delta V$ al escribir que el trabajo realizado por la batería es $E\Delta Z$. Pero resulta que es muy difícil técnicamente mantener el volumen constante. Es mucho más fácil mantener la celda a presión atmosférica constante. Es por ello que a los químicos no les gusta ninguna de las ecuaciones que hemos escrito más arriba; prefieren ecuaciones que describen el comportamiento a *presión* constante. Al principio de este capítulo elegimos V y T como variables independientes. Los químicos prefieren P y T, por lo que ahora consideraremos cómo se pueden transformar los resultados obtenidos hasta aquí al sistema de variables de los químicos. Recuerden que en el tratamiento que sigue es fácil de que haya confusión debido al cambio de engranaje de T y V a T y P.

Partimos de la ecuación (45.3) con $\Delta U = \Delta Q - P\Delta V$; se puede remplazar $P\Delta V$ por $E\Delta Z$ o $A\Delta B$. Si de alguna manera pudiéramos remplazar el último término, $P\Delta V$, por $V\Delta P$, habríamos intercambiado V y P y los químicos se pondrían contentos. Bien, un hombre inteligente notó que el diferencial del producto PV es $d(PV) = P\,dV + V\,dP$ y que si sumaba esta igualdad a la ecuación (45.3) obtenía

$$\begin{array}{rl} \Delta(PV) = & P\,\Delta V + V\,\Delta P \\ \Delta U = & \Delta Q \quad - P\,\Delta V \\ \hline \Delta(U+PV) = & \Delta Q \quad + V\,\Delta P \end{array}$$

Para que el resultado se parezca a la ecuación (45.3) definimos $U + PV$ como algo nuevo: la *entalpía*, H, y escribimos $\Delta H = \Delta Q + V\Delta P$.

Ahora estamos listos para traducir nuestros resultados al lenguaje de los químicos con las siguientes reglas: $U \to H, P \to -V, V \to P$. Por ejemplo, la relación fundamental que usarían los químicos en vez de la ecuación (45.7) es:

$$\left(\frac{\partial H}{\partial P}\right)_T = -T\left(\frac{\partial V}{\partial T}\right)_P + V.$$

Ahora debe estar claro cómo se transforma a las variables T y P de los químicos. Volvamos a nuestras variables originales, en lo que queda del capítulo T y V serán las variables independientes.

Apliquemos ahora los resultados obtenidos a un cierto número de situaciones físicas. Consideremos primero el gas ideal. Por la teoría cinética sabemos que la energía interna de un gas sólo depende del movimiento de las moléculas y del número de éstas. La energía interna depende de T pero no de V. Si cambiamos

V manteniendo T constante, U no cambia. En consecuencia $(\partial U/\partial V)_T = 0$ y la ecuación (45.7) nos dice que para un gas ideal,

$$T\left(\frac{\partial P}{\partial T}\right)_V - P = 0. \tag{45.10}$$

La ecuación (45.10) es una ecuación diferencial que nos puede decir algo acerca de P. Consideremos la derivada parcial del siguiente modo: como la derivada parcial es a volumen constante, la remplazaremos por una derivada ordinaria y para acordarnos escribiremos explícitamente "V constante". Entonces, la ecuación (45.10) se convierte en:

$$T\,\frac{\Delta P}{\Delta T} - P = 0; \qquad \text{constante } V, \tag{45.11}$$

que podemos integrar obteniendo:

$$\begin{aligned} \ln P &= \ln T + \text{constante}; \qquad &\text{constante } V,\\ P &= \text{constante} \times T; \qquad &\text{constante } V. \end{aligned} \tag{45.12}$$

Sabemos que para un gas ideal la presión por mol es igual a:

$$P = \frac{RT}{V}, \tag{45.13}$$

que es compatible con (45.12) puesto que V y R son constantes. ¿Para qué nos tomamos el trabajo de hacer este cálculo si ya sabíamos los resultados? ¡Porque hemos estado usando *dos definiciones independientes de temperatura*! En un cierto momento supusimos que la energía cinética de las moléculas era proporcional a la temperatura, suposición que define una escala de temperatura que llamaremos escala del gas ideal. La T de la ecuación (45.13) se basa en esta escala. También llamamos temperaturas *cinéticas* a las temperaturas basadas en la escala de gases. Más tarde definimos la temperatura de un segundo modo completamente independiente de cualquier sustancia. A partir de razonamientos basados en la segunda ley definimos lo que podríamos llamar "gran temperatura termodinámica absoluta" T; es la T que aparece en la ecuación (45.12). Lo que hemos demostrado aquí es que la presión de un gas ideal (definido como aquel en que la energía interna no depende del volumen) es proporcional a la gran temperatura termodinámica absoluta. También sabemos que la presión es proporcional a la temperatura medida en la escala de gases. En consecuencia, podemos deducir que la temperatura cinética es proporcional a la "gran temperatura termodinámica absoluta". Esto significa, naturalmente, que si fuésemos sensatos podríamos hacer concordar las dos escalas. En este caso, al menos, las dos escalas *han sido* elegidas de modo que coincidan; se ha escogido 1 para la constante de proporcionalidad. La mayoría de las veces el hombre se busca las dificultades, pero en este caso ¡ha hecho las dos escalas iguales!

45-3 La ecuación de Clausius-Clapeyron

La vaporización de un líquido es otra aplicación de los resultados que hemos obtenido. Supongan que tenemos un líquido en un cilindro, de modo que lo podemos comprimir empujando el pistón, y que nos preguntamos: "Si mantenemos la temperatura constante, ¿cómo varía la presión con el volumen?" Es decir, queremos dibujar una isoterma en el diagrama P-V. La sustancia que hay en el cilindro no es el gas ideal que consideramos antes; ahora puede que esté en la fase líquida o en la fase vapor, o ambas pueden estar presentes. Si aplicamos presión suficiente, la sustancia se condensa y se vuelve líquida. Si ahora apretamos aún más fuerte, el volumen cambia muy poco y la isoterma sube rápidamente al disminuir el volumen, como muestra la parte izquierda de la figura (45.3).

Si aumentamos el volumen tirando del pistón, la presión cae hasta que alcanzamos el punto en que el líquido comienza a hervir y comienza a formarse vapor. Si tiramos aún más del pistón, todo lo que ocurre es que se evapora más

líquido. Cuando hay en parte líquido y en parte vapor en el cilindro, las dos fases están en equilibrio —el líquido se evapora a la misma velocidad que el gas se condensa—. Si hacemos más espacio para el vapor, se necesita más vapor para mantener la presión, por lo que se evapora un poco más de líquido, pero la presión se mantiene constante. En la parte horizontal de la curva de la figura 45-3 la presión no cambia; este valor constante se denomina *presión de vapor a la temperatura* T. A medida que seguimos aumentando el volumen se llega a un punto en que ya no hay más líquido para evaporar. En este punto, si expandimos el volumen aún más, la presión bajará como en un gas ordinario, como muestra la parte derecha del diagrama P-V. La curva inferior de la figura 45-3 es la isoterma correspondiente a una temperatura $T - \Delta T$ ligeramente menor. La presión de la fase líquida se reduce ligeramente porque el líquido se expande al aumentar la temperatura (para la mayoría de las sustancias, pero no para el agua cerca del punto de congelación) y naturalmente la presión de vapor es menor a la temperatura menor.

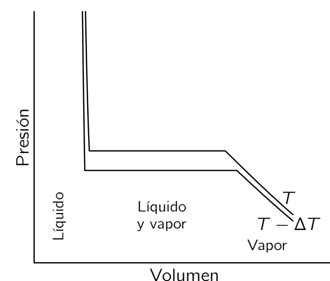


Figura 45-3. Isotermas de un vapor condensable en un cilindro. A la izquierda la sustancia está en la fase líquida. A la derecha la sustancia está vaporizada. Al centro, líquido y vapor coexisten en el cilindro.

Ahora haremos un ciclo con las dos isotermas uniéndolas (digamos por líneas adiabáticas) por ambos extremos de las secciones horizontales superiores, como muestra la figura 45-4. La pequeña oscilación en el extremo inferior derecho de la figura no altera prácticamente nada y la despreciaremos. Vamos a usar el razonamiento de Carnot que plantea que el calor agregado a la sustancia para hacerla pasar de líquido a vapor está relacionado con el trabajo efectuado por la sustancia, al recorrer el ciclo. Además, llamemos L al calor necesario para vaporizar la sustancia contenida en el cilindro. Por el razonamiento inmediato anterior a la ecuación (45.5) sabemos que $L(\Delta T/T) =$ trabajo realizado por la sustancia. Como antes, el trabajo realizado por la sustancia es el área sombreada, que es aproximadamente $\Delta P(V_G - V_L)$, donde ΔP es la diferencia de presión de vapor a las dos temperaturas T y $T - \Delta T$, V_G es el volumen del gas y V_L el del líquido, ambos medidos a la presión de vapor a temperatura T. Igualando estas dos expresiones del área, obtenemos $L\,\Delta T/T = \Delta P(V_G - V_L)$, es decir,

$$\frac{L}{T(V_G - V_L)} = (\partial P_{\text{vap}}/\partial T). \tag{45.14}$$

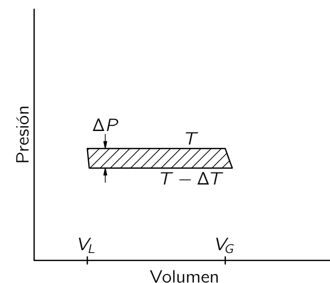


Figura 45-4. Diagrama presión-volumen para un ciclo de Carnot con un vapor condensable en el cilindro. A la izquierda, la sustancia está en estado líquido. Se agrega una cantidad de calor L a temperatura T para evaporar el líquido. El vapor se expande, adiabáticamente cuando T pasa a $T - \Delta T$.

La ecuación (45.14) da la relación entre la rapidez de variación de la presión de vapor con la temperatura y la cantidad de calor necesaria para evaporar el líquido. Esta relación fue deducida por Carnot pero se llama ecuación de Clausius-Clapeyron.

Comparemos ahora la ecuación (45.14) con los resultados deducidos de la teoría cinética. Por lo común, V_G es mucho mayor que V_L. Así, pues, $V_G - V_L \approx V_G = RT/P$ por mol. Si además suponemos que L es constante, independiente de la temperatura —no es muy buena aproximación— tendremos $\partial P/\partial T = L/(RT^2/P)$. La solución de esta ecuación diferencial es:

$$P = \text{constante}\, e^{-L/RT}. \tag{45.15}$$

Comparemos ésta con la variación de presión de la temperatura, deducida anteriormente en la teoría cinética. Al menos aproximadamente, la teoría cinética indicaba la posibilidad de que el número de moléculas de vapor por unidad de volumen encima del líquido fuese:

$$n = \left(\frac{1}{V_a}\right) e^{-(U_G - U_L)/RT}, \tag{45.16}$$

donde $U_G - U_L$ es la energía interna por mol del gas menos la energía interna por mol del líquido, es decir, la energía necesaria para vaporizar un mol de líquido. La ecuación (45.15) proveniente de la termodinámica, y la ecuación (45.16), proveniente de la teoría cinética, están estrechamente relacionadas porque la presión es nkT, pero no son exactamente iguales. Sin embargo, resultarán exactamente iguales si suponemos $U_G - U_L =$ constante en vez de $L =$ constante. Si suponemos $U_G - U_L =$ constante independiente de la temperatura, el razonamiento que nos llevó a la ecuación (45.15) también llevará a la ecuación (45.16). Como

la presión es constante mientras el volumen está cambiando, el cambio en la energía interna $U_G - U_L$ es igual al calor L puesto menos el trabajo realizado $P(V_G - V_L)$, entonces $L = (U_G + PV_G) - (U_L + PV_L)$.

La comparación muestra la ventajas y desventajas de la termodinámica sobre la teoría cinética: primero, la ecuación (45.14), obtenida termodinámicamente es exacta, mientras que la ecuación (45.16) sólo se puede aproximar, por ejemplo, si U es casi constante y si el modelo es válido. Segundo, no podemos comprender correctamente cómo es que el gas pasa al estado líquido; no obstante, la ecuación (45.14) es exacta mientras que la (45.16) es sólo aproximada. Tercero, aunque nuestro tratamiento se aplica a un gas que se condensa en un líquido, el razonamiento es válido para cualquier otro cambio de estado. Por ejemplo, la transición sólido líquido tiene el mismo tipo de curva que el mostrado en la figura 45-3 o 45-4. Introduciendo el calor latente de fusión, M/mol, la fórmula análoga a la ecuación (45.14), es $(\partial P_{\text{fus}}/\partial T)_V = M/[T(V_{\text{líq}} - V_{\text{sol}})]$. Aunque no podamos comprender la teoría cinética del proceso de fusión tenemos una ecuación exacta. Sin embargo, cuando *podemos* comprender la teoría cinética tenemos otra ventaja. La ecuación (45.14) es únicamente una relación diferencial y no hay manera de determinar las constantes de integración. En la teoría cinética podemos obtener las constantes también si tenemos un buen modelo que describa completamente el fenómeno. Cada una tiene, pues, ventajas y desventajas. Cuando el conocimiento es escaso y la situación complicada, las relaciones termodinámicas son realmente las más poderosas. Cuando la situación es muy simple y se puede hacer un análisis teórico, es mejor tratar de obtener más información con el análisis teórico.

Un ejemplo más: la radiación de cuerpo negro. Hemos estudiado una caja que contiene radiación y nada más. Hemos hablado del equilibrio entre el oscilador y la radiación. También encontramos que los fotones que chocan contra las paredes de la caja deben ejercer una presión P y hallamos $PV = U/3$, donde U es la energía total de todos los fotones y V es el volumen de la caja. Si introducimos $U = 3PV$ en la ecuación básica (45.7), encontramos*

$$\left(\frac{\partial U}{\partial V}\right)_T = 3P = T\left(\frac{\partial P}{\partial T}\right)_V - P. \tag{45.17}$$

Como el volumen de nuestra caja es constante, podemos remplazar $(\partial P/\partial T)_V$ por dP/dT obteniendo una ecuación diferencial ordinaria que podemos integrar: $\ln P = 4 \ln T + \text{constante}$, o sea $P = \text{constante} \times T^4$. La presión de radiación varía como la cuarta potencia de la temperatura y la densidad de energía de la radiación, $U/V = 3P$, también varía como T^4. Se acostumbra escribir $U/V = (4\sigma/c)T^4$, donde c es la velocidad de la luz y σ es llamada *constante de Stefan-Boltzmann.* No es posible obtener σ únicamente con la termodinámica. Éste es un buen ejemplo de su fuerza y su limitación. Saber que U/V varía como T^4 ya es mucho, pero saber cuánto vale verdaderamente U/V a cualquier temperatura requiere que entremos en un tipo de detalles que sólo una teoría completa nos puede dar. Tenemos esa teoría para la radiación de cuerpo negro y podemos obtener una expresión para la constante σ de la siguiente manera.

Sea $I(\omega)\,d\omega$ la distribución de intensidad, o sea la energía que fluye a través de 1 m^2 en un segundo con frecuencia entre ω y $\omega + d\omega$. La distribución de densidad de energía = energía/volumen = $I(\omega)\,d\omega/c$ es:

$$\begin{aligned}\frac{U}{V} &= \text{densidad de energía total}\\ &= \int_{\omega=0}^{\infty} \text{densidad de energía entre } \omega \text{ y } \omega + d\omega\\ &= \int_0^{\infty} \frac{I(\omega)\,d\omega}{c}.\end{aligned}$$

* En este caso $(\partial P/\partial V)_T = 0$, porque para mantener el oscilador en equilibrio a una temperatura dada, la radiación en la vecindad del oscilador tiene que ser la misma, independientemente del volumen de la caja. Por lo tanto, la cantidad total de fotones dentro de la caja debe ser proporcional a su volumen, por lo que la energía interna por unidad de volumen y, por lo tanto, la presión, depende únicamente de la temperatura.

Por nuestras discusiones anteriores sabemos que

$$I(\omega)=\frac{\hbar\omega^3}{\pi^2c^2(e^{\hbar\omega/kT}-1)}.$$

Sustituyendo esta expresión de $I(\omega)$ en nuestra ecuación para U/V obtenemos:

$$\frac{U}{V}=\frac{1}{\pi^2c^3}\int_0^\infty\frac{\hbar\omega^3\,d\omega}{e^{\hbar\omega/kT}-1}.$$

Si ponemos $x=\hbar\omega/kT$, la expresión se transforma en

$$\frac{U}{V}=\frac{(kT)^4}{\hbar^3\pi^2c^3}\int_0^\infty\frac{x^3\,dx}{e^x-1}.$$

Esta integral es simplemente un número que podemos obtener aproximadamente dibujando una curva y calculando el área contando cuadrados. Vale 6.5 más o menos. Los matemáticos que estén entre nosotros pueden demostrar que la integral es exactamente $\pi^4/15$.* Comparando esta expresión con $U/V=(4\sigma/c)T^4$ encontramos:

$$\sigma=\frac{k^4\pi^2}{60\hbar^3c^3}=5.67\times10^{-8}\,\frac{\text{vatio}}{(\text{metro})^2(\text{grado})^4}.$$

Si hacemos un pequeño agujero en nuestra caja, ¿cuánta energía saldrá por segundo por el agujero de área unitaria? Para pasar de la densidad de energía al flujo de energía, multiplicamos la densidad de energía U/V por c. También multiplicamos por $\frac{1}{4}$, que proviene de lo que sigue: primero, un factor $\frac{1}{2}$ porque sólo escapa la energía que fluye hacia afuera del *agujero*, y segundo, otro factor $\frac{1}{2}$ porque la energía que se acerca al agujero en dirección oblicua tiene menos posibilidades de salir por estar afectada con un factor coseno. El valor medio del factor coseno es $\frac{1}{2}$. Ahora se ve claro por qué escribimos $U/V=(4\sigma/c)T^4$, para que al final podamos decir que el flujo que proviene de un pequeño agujero es σT^4 por unidad de área.

* Como $(e^x-1)^{-1}=e^{-x}+e^{-2x}+\cdots$, la integral es

$$\sum_{n=1}^{\infty}\int_0^\infty e^{-nx}x^3\,dx.$$

Pero $\int_0^\infty e^{-nx}dx=1/n$, que derivada tres veces respecto a n da $\int_0^\infty x^3e^{-nx}dx=6/n^4$, por lo que la integral es $6(1+\frac{1}{16}+\frac{1}{18}+\cdots)$ y se obtiene una buena aproximación sumando los primeros términos. En el capítulo 50 encontraremos una manera de demostrar que la suma de las inversas de la cuarta potencia de los enteros es, en efecto, $\pi^4/90$.

46

Rueda dentada y trinquete*

46-1 Cómo trabaja una rueda dentada

En este capítulo discutiremos la rueda dentada y trinquete, un dispositivo muy simple que permite que un eje gire solamente en un sentido. La posibilidad de que tengamos algo que gire solamente en un sentido requiere un cierto análisis detallado y cuidadoso, hay algunas consecuencias muy interesantes.

El plan de la discusión surgió cuando se trató de inventar una explicación elemental, desde el punto de vista molecular o cinético, para el hecho de que hay una cantidad máxima de trabajo que se puede extraer de una máquina térmica. Naturalmente, hemos visto la esencia del razonamiento de Carnot, pero sería bueno encontrar una explicación que sea elemental en el sentido de que podamos ver lo que está sucediendo físicamente. Ahora bien, hay demostraciones matemáticas complicadas tomando como base las leyes de Newton para demostrar que sólo podemos obtener una cierta cantidad de trabajo cuando el calor fluye de un lugar a otro, pero es muy difícil hacer esto en una demostración elemental. Abreviando, no lo entendemos aunque podamos seguir el proceso matemático.

En el razonamiento de Carnot, el hecho de que no se pueda extraer más que una cierta cantidad de trabajo al ir de una temperatura a otra se deduce de otro axioma: si todo está a la misma temperatura, el calor no se puede convertir en trabajo mediante un proceso cíclico. En primer lugar, volvamos atrás y tratemos de ver, al menos en un ejemplo elemental, por qué es verdadera esta afirmación tan simple.

Tratemos de inventar un dispositivo que viole la Segunda ley de la termodinámica, esto es, un artefacto que saque trabajo de un depósito de calor, estando todo a la misma temperatura. Digamos que tenemos una caja de gas a una cierta temperatura y que dentro hay un eje con aspas. (véase figura 46-1, pero tomen $T_1 = T_2 = T$, por ejemplo.) A causa del bombardeo de las moléculas de gas sobre las aspas, éstas oscilan y se mueven con pequeñas sacudidas. Todo lo que tenemos que hacer es colocar en el otro extremo del eje una rueda que sólo pueda moverse en un sentido —rueda dentada y trinquete—. Así, cuando el eje trate de vibrar en un sentido, no girará, y cuando vibre en el otro, sí girará. La rueda girará lentamente, y ¡quizás hasta podríamos atar una pulga a una cuerda que pende de un tambor en el eje y levantarla! Preguntémonos si esto es posible. Según la hipótesis de Carnot es imposible. Pero si la miramos, a *prima facie* parece muy posible. Por lo tanto, tenemos que considerarlo más cuidadosamente. Verdaderamente, si consideramos la rueda dentada y trinquete vemos muchas complicaciones.

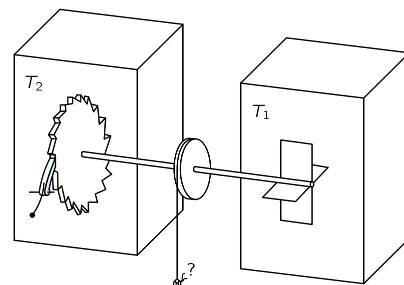


Figura 46-1. La máquina dentada-trinquete rueda.

En primer lugar, nuestra rueda dentada idealizada es lo más simple posible, pero aun así hay un trinquete y éste debe tener un resorte. El trinquete debe volver después de pasar un diente, por lo que es necesario el resorte.

Otro rasgo de esta rueda dentada y trinquete, que la figura no muestra, es completamente esencial. Supongan que el dispositivo estuviese formado de partes perfectamente elásticas. Después que el trinquete ha sido levantado hasta el final del diente y es empujado por el resorte, rebotará contra la rueda y continuará rebotando. Entonces, cuando llegue otra fluctuación, la rueda podría girar en el otro sentido, ¡ya que el diente podría pasar por debajo durante el momento en que el trinquete estuvo levantado! Por lo cual, una parte esencial de la irreversibilidad de nuestra rueda es un mecanismo amortiguador que pare los rebotes. Cuando

* Véase Parrondo y Español, *Am. J. Phys.* **64**, 1125 (1996) para un análisis crítico del capítulo.

hay amortiguamiento, naturalmente, la energía que había en el trinquete pasa a la rueda y se transforma en calor. De modo que la rueda se irá calentando a medida que gira. Para simplificar las cosas, podemos colocar un gas alrededor de la rueda para que tome parte del calor. De todas formas, digamos que el gas aumenta en temperatura al mismo tiempo que la rueda. ¿Seguirá esto así para siempre? ¡No! El trinquete y la rueda, ambos a cierta temperatura T, también tienen movimiento browniano. Este movimiento es tal que de vez en cuando, accidentalmente, el trinquete se levanta sobre un diente exactamente en el momento en que el movimiento browniano de las aspas está tratando de girar el eje hacia atrás. Y cuanto más se calienta todo, más frecuentemente ocurre esto.

Por lo tanto, ésta es la razón por la que este dispositivo no trabaja en movimiento perpetuo. Cuando las aspas son empujadas, algunas veces el trinquete se levanta sobre un diente. Pero otras veces, cuando trata de girar en el otro sentido, el trinquete se ha levantado ya debido a las fluctuaciones de los movimientos en el extremo de la rueda ¡y ésta se vuelve en el otro sentido! El resultado neto es nada. No es difícil demostrar que cuando la temperatura en los dos lados es igual, no habrá movimiento promedio neto de la rueda. Naturalmente, la rueda vibrará mucho en un sentido y en el otro, pero no hará lo que nosotros queríamos, que es girar sólo en un sentido.

Veamos la razón. Es necesario realizar trabajo contra el resorte para que se levante el trinquete hasta el tope de un diente. Llamemos ϵ a esta energía y sea θ el ángulo entre dientes. La probabilidad de que el sistema pueda acumular suficiente energía ϵ para llevar el trinquete hasta el tope de un diente es $e^{-\epsilon/kT}$. Pero la probabilidad de que el trinquete esté accidentalmente levantado es también $e^{-\epsilon/kT}$. Por lo cual el número de veces que el trinquete está levantado y la rueda puede girar hacia atrás libremente es igual al número de veces que hay suficiente energía para girarla hacia adelante cuando el trinquete está abajo. Así obtenemos un "equilibrio" y la rueda no girará.

46-2 La rueda dentada como máquina

Demos ahora un paso más. Tomemos como ejemplo cuando la temperatura de las aspas es T_1 y la temperatura de la rueda, rueda dentada, es T_2 y que T_2 es menor que T_1. Como la rueda está fría y las fluctuaciones del trinquete no son relativamente frecuentes, será muy difícil que el trinquete alcance una energía ϵ. A causa de la alta temperatura T_1, las aspas alcanzarán con frecuencia la energía ϵ, por lo que nuestro dispositivo irá en una dirección, como se diseñó.

Ahora querríamos ver si podemos levantar pesos. Atamos una cuerda al tambor del medio y colocamos un peso en la cuerda, como nuestra pulga, por ejemplo. Digamos que L es el torque debido al peso. Si L no es muy grande, nuestra máquina levantará el peso porque las fluctuaciones brownianas hacen que se mueva con más probabilidad en un sentido que en otro. Queremos encontrar qué peso puede levantar, con qué rapidez gira, etcétera.

Consideremos, en primer lugar, un movimiento hacia adelante, que es el modo corriente en que se diseña una rueda dentada para que gire. Para que dé un paso hacia adelante, ¿cuánta energía se tiene que tomar prestada de las aspas? Tenemos que tomar prestada una energía ϵ para levantar el trinquete. La rueda gira un ángulo θ contra el torque L, por lo que también necesitamos la energía $L\theta$. La cantidad total de energía que tenemos que tomar es, pues, $\epsilon + L\theta$. La probabilidad de que la obtengamos es proporcional a $e^{-(\epsilon+L\theta)/kT_1}$. En realidad, no es sólo cuestión de obtener la energía, sino que también querríamos saber el número de veces por segundo que tiene esta energía. La probabilidad por segundo es proporcional a $e^{-(\epsilon+L\theta)/kT_1}$, y llamaremos $1/\tau$ a la constante de proporcionalidad. De todos modos se cancelará al final. Cuando sucede un paso hacia adelante, el trabajo efectuado sobre el peso es $L\theta$. La energía tomada de las aspas es $\epsilon + L\theta$. El resorte se pone tenso con energía ϵ y luego, cataplún, se dispara y esta energía se transforma en calor. Toda la energía que se ha tomado

Tabla 46-1

Resumen del funcionamiento de la rueda dentada y trinquete.

Hacia adelante:	Necesita la energía	$\epsilon + L\theta$	de las aspas. $\therefore$ Probabilidad por segundo $= \dfrac{1}{\tau}\, e^{-(L\theta+\epsilon)/kT_1}$
	Toma de las aspas	$L\theta + \epsilon$	
	Efectúa el trabajo	$L\theta$	
	Da a la rueda dentada	ϵ	
Hacia atrás:	Necesita la energía	ϵ	para el trinquete. $\therefore$ Probabilidad por segundo $= \dfrac{1}{\tau}\, e^{-\epsilon/kT_2}$
	Toma de la rueda	ϵ	lo mismo que antes con signos opuestos.
	Libera el trabajo	$L\theta$	
	Da a las aspas	$L\theta + \epsilon$	

Si el sistema es reversible, las probabilidades por segundo son iguales, por lo que $\dfrac{\epsilon + L\theta}{T_1} = \dfrac{\epsilon}{T_2}$.

$\dfrac{\text{Calor con la rueda dentada}}{\text{Calor de las aspas}} = \dfrac{\epsilon}{L\theta + \epsilon}$. De aquí $\dfrac{Q_2}{Q_1} = \dfrac{T_2}{T_1}$.

se emplea en levantar el peso y en accionar el trinquete, que luego cae y da calor al otro lado del diente.

Consideremos ahora el caso contrario que es el movimiento hacia atrás. ¿Qué sucede aquí? Para hacer que la rueda vaya hacia atrás todo lo que tenemos que hacer es dar energía para levantar el trinquete lo bastante alto para que la rueda se escape. Ésta es aún la energía ϵ. La probabilidad por segundo de que el trinquete se levante esa altura es ahora $(1/\tau)e^{-\epsilon/kT_2}$. La constante de proporcionalidad es la misma, pero ahora aparece kT_2 porque la temperatura es diferente. Cuando esto sucede, se libera trabajo, ya que la rueda va hacia atrás. Pierde una muesca y libera el trabajo $L\theta$. La energía tomada del sistema dentado es ϵ, y la energía dada al gas a T_1, donde están las aspas, es $L\theta + \epsilon$. Hay que pensar un poco para ver la razón de esto. Supongan que el trinquete se ha levantado por sí mismo accidentalmente debido a una fluctuación. Entonces, cuando vuelve a caer y el resorte lo empuja contra el diente, hay una fuerza que trata de girar la rueda porque el diente está empujando en un plano oblicuo. Esta fuerza realiza trabajo y lo mismo hace la fuerza debida a los pesos. Por lo cual las dos juntas forman la fuerza total y toda la energía que se va liberando lentamente aparece en el extremo con aspas en forma de calor. (Naturalmente debe ser así por la conservación de la energía, ¡pero se debe tener cuidado de considerar las cosas en su totalidad!) Vemos que todas estas energías son exactamente iguales pero invertidas. Por lo que, según cuál de estas dos probabilidades por segundo sea mayor, el peso será levantado lentamente o dejado caer despacio. Naturalmente estará sacudiéndose de forma constante, yendo hacia arriba durante un instante y hacia abajo durante otro, pero estamos hablando del comportamiento medio.

Supongan que para un peso particular sucede que las probabilidades son iguales. Entonces, añadimos un peso infinitesimal a la cuerda. El peso descenderá lentamente y se efectuará trabajo en la máquina. Se tomará energía de la rueda y se lo dará a las aspas. Si por el contrario quitamos un poquito de peso, el desajuste ocurrirá al revés. El peso es levantado, y se toma calor de las aspas y se lo da a la rueda. Por consiguiente, tenemos las condiciones de un ciclo reversible de Carnot, siempre que el peso sea tal que las dos probabilidades son iguales. Esta condición es evidentemente que $(\epsilon + L\theta)/T_1 = \epsilon/T_2$. Digamos que la máquina está levantando lentamente el peso. Se toma una energía Q_1 de las aspas y se entrega una energía Q_2 a la rueda, y las dos están en la proporción $(\epsilon + L\theta)/\epsilon$. Si estamos bajando el peso, también tenemos $Q_1/Q_2 = (\epsilon + L\theta)/\epsilon$.

Entonces (tabla 46-1) tenemos:

$$Q_1/Q_2 = T_1/T_2.$$

Aún más, el trabajo que obtenemos es a la energía tomada de las aspas como $L\theta$ es a $L\theta + \epsilon$, y por lo tanto como $(T_1 - T_2)/T_1$. Vemos que nuestro artefacto no puede extraer más trabajo que éste, operando reversiblemente. Éste es el resultado que esperábamos del razonamiento de Carnot y el principal resultado de esta lección. Sin embargo, podemos usar nuestro artefacto para entender algunos otros fenómenos, aunque sea fuera del equilibrio, y por lo tanto más allá del ámbito de la termodinámica.

Calculemos ahora *con qué rapidez* giraría nuestro artefacto de un solo sentido si todo estuviese a la misma temperatura y colgásemos un peso del tambor. Si tiramos muy, pero muy fuerte, naturalmente hay toda clase de complicaciones. El trinquete se desliza por encima de la rueda, o el resorte se rompe, o cualquier otra cosa. Pero supongan que tiramos lo suficientemente suave como para que todo trabaje perfectamente. En estas circunstancias, el análisis anterior es correcto para la probabilidad de que la rueda vaya hacia adelante o hacia atrás, si recordamos que las dos temperaturas son iguales. En cada paso se obtiene un ángulo θ, por lo que la velocidad angular es θ por la probabilidad de uno de estos saltos por segundo. Va hacia adelante con una probabilidad $(1/\tau)e^{-(\epsilon+L\theta)/kT}$ y hacia atrás con una probabilidad $(1/\tau)e^{-\epsilon/kT}$, y obtenemos para la velocidad angular

$$\begin{aligned} \omega &= (\theta/\tau)(e^{-(\epsilon+L\theta)/kT} - e^{-\epsilon/kT}) \\ &= (\theta/\tau)e^{-\epsilon/kT}(e^{-L\theta/kT} - 1). \end{aligned} \tag{46.1}$$

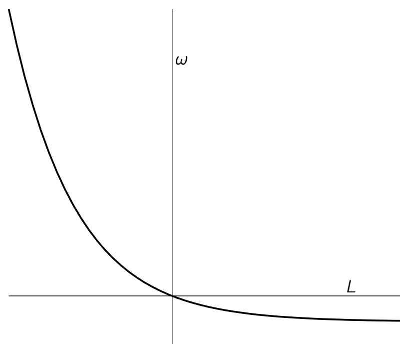


Figura 46-2. Velocidad angular de la rueda dentada en función del torque.

Si representamos ω en función de L, obtenemos la curva que muestra la figura 46-2. Vemos que es muy distinto que L sea positiva o negativa. Si L aumenta positivamente, lo que sucede cuando tratamos de llevar la rueda hacia atrás, la velocidad hacia atrás se aproxima a una constante. Cuando L se hace negativa, ω realmente "despega", ya que ¡e a una enorme potencia es muy grande!

La velocidad angular que se obtuvo de fuerzas diferentes es, por consiguiente, muy asimétrica. Si vamos en un sentido es muy fácil: tenemos una velocidad angular grande para una fuerza pequeña. Yendo en el otro, podemos aplicar una fuerza muy grande y aún así apenas gira la rueda.

Encontramos lo mismo en un *rectificador eléctrico*. En lugar de la fuerza tenemos un campo eléctrico y en lugar de la velocidad angular, una corriente eléctrica. En el caso de un rectificador, el voltaje no es proporcional a la resistencia y la situación es asimétrica. El mismo análisis que hicimos para el rectificador mecánico, servirá también para un rectificador eléctrico. De hecho, la fórmula que acabamos de obtener es típica de la capacidad de los rectificadores para conducir corriente en función de sus voltajes.

Quitemos ahora todos los pesos y consideremos la máquina original. Si T_2 fuese menor que T_1, la rueda dentada iría hacia adelante, cosa que cualquiera está dispuesto a creer. Pero lo que es difícil de creer, a primera vista, es lo contrario. Si T_2 es mayor que T_1, ¡la rueda gira en sentido opuesto! Una rueda dentada dinámica con un montón de calor gira por sí misma hacia atrás, ya que su trinquete está rebotando. Si el trinquete está inclinado por un momento en algún punto, empuja el plano inclinado lateralmente. Pero *siempre* está empujando en un plano inclinado, porque si sucede que es levantado lo suficientemente alto para que pase el tope de un diente, entonces el plano inclinado se desliza y vuelve a caer en un plano inclinado. Por consiguiente, ¡una rueda dentada y su trinquete, calientes, están idealmente construidos para girar en una dirección exactamente opuesta a aquella para la cual fueron diseñados originalmente!

A pesar de toda nuestra maestría en diseños desequilibrados, si las dos temperaturas son exactamente iguales, no hay mayor propensión a girar en un sentido que en otro. En el momento en que estemos observando puede que esté girando en un sentido o en otro, pero a lo largo de la carrera no llega a ningún

sitio. Este hecho de que no llega a ningún sitio es realmente el profundo principio fundamental en el que se basa toda la termodinámica.

46-3 Reversibilidad en mecánica

¿Qué principio mecánico más profundo nos dice que, a la larga, si se conserva la temperatura igual en todas partes, nuestro artefacto no girará ni a la derecha ni a la izquierda? Tenemos evidentemente una proposición fundamental: no hay modo de diseñar una máquina que, abandonada a sí misma, con más probabilidad de girar en un sentido que en otro después de un tiempo suficientemente largo. Debemos ver cómo esto es una consecuencia de las leyes de la mecánica.

Las leyes de la mecánica son más o menos así: la masa por la aceleración es la fuerza, y la fuerza sobre una partícula es una función complicada de las posiciones de todas las otras partículas. Hay otras situaciones en las que la fuerza depende de la velocidad, como en el magnetismo, pero no consideremos esto ahora. Tomemos un caso más sencillo, como la gravedad, donde las fuerzas dependen únicamente de la posición. Supongan, pues, que hemos resuelto nuestro conjunto de ecuaciones y obtenido un cierto movimiento $x(t)$ para cada partícula. En un sistema lo suficientemente complicado, las soluciones son muy complicadas y lo que sucede en el tiempo resulta muy sorprendente. Si escribimos cualquier arreglo que deseemos para las partículas, ¡veremos que este arreglo ocurre realmente si esperamos un tiempo suficientemente largo! Si seguimos nuestra solución durante un tiempo suficientemente largo, ella intenta hacer todo lo que puede, por así decir. Esto no es absolutamente necesario en los dispositivos más simples, pero cuando los sistemas se hacen lo suficientemente complicados, con átomos suficientes, sucede así. La solución también puede hacer algo más. Si resolvemos las ecuaciones de movimiento, podemos obtener ciertas funciones como $t + t^2 + t^3$. Sostenemos que otra solución también sería $-t + t^2 - t^3$. En otras palabras, si sustituimos t en todas partes por $-t$ en la solución completa, también obtendremos una solución de la misma ecuación. Esto es una consecuencia del hecho de que si sustituimos t por $-t$ en la ecuación diferencial original, nada se cambia, ya que aparecen solamente las segundas derivadas respecto a t. Esto significa que si tenemos un cierto movimiento, el movimiento exactamente opuesto también es posible. En la confusión total que resulta de esperar un tiempo suficientemente largo, se encuentra que unas veces va en una dirección y otras, en la dirección opuesta. Ninguno de los dos movimientos es mejor que el otro. Por lo tanto, es imposible diseñar una máquina que, a la larga, vaya con más probabilidad en una dirección que en la opuesta, si la máquina es lo suficientemente complicada.

Uno podría pensar en un ejemplo para el cual esto es evidentemente falso. Si tomamos una rueda, por ejemplo, y hacemos que gire en el espacio vacío, lo hará en el mismo sentido para siempre. Por ello, hay algunas condiciones, como la conservación del momentum angular, que violan el razonamiento anterior. Además, esto requiere solamente de que el razonamiento se haga con un poquito más de cuidado. Quizá las paredes toman el momentum angular, o algo parecido, y de este modo no tenemos leyes de conservación especiales. Entonces, si el sistema es lo necesariamente complicado, el razonamiento es válido. Se basa en el hecho de que las leyes de la mecánica son reversibles.

Por interés histórico querríamos señalar un dispositivo inventado por Maxwell, que fue el primero en desarrollar la teoría dinámica de los gases. Supuso la situación siguiente: tenemos dos cajas de gas a la misma temperatura, con un pequeño agujero entre ellas. En el agujero se sienta un pequeño demonio (¡que puede ser una máquina, naturalmente!). El agujero tiene una puerta que el demonio puede abrir o cerrar. Él observa las moléculas que vienen de la izquierda. Siempre que ve una molécula rápida abre la puerta. Cuando ve una lenta deja la puerta cerrada. Si queremos que sea un demonio de primera, puede tener ojos en la nuca, y hacer lo contrario con las moléculas de la derecha. Deja que pasen las lentas a la izquierda y las rápidas a la derecha. Muy pronto la parte izquierda estará fría y la derecha caliente. Ahora bien, ¿se han violado las ideas de la termodinámica porque podemos tener dicho demonio?

Resulta que, si construimos un demonio de tamaño finito, el mismo demonio se pondrá caliente después de un momento que no podrá ver claro. El demonio más sencillo posible sería, por ejemplo, una puerta trampa sujeta a la parte superior del agujero mediante un resorte. Una molécula rápida atraviesa porque es capaz de levantar la puerta trampa. La molécula lenta no puede atravesar y rebota hacia atrás. Pero esto no es otra cosa que nuestra rueda y trinquete un poco cambiados, y finalmente el mecanismo se calentará. Si suponemos que el calor específico del demonio no es infinito, se debe calentar. Tiene solamente un número finito de engranajes y ruedas internos, por lo que no se puede desembarazar del calor adicional que obtiene de la observación de las moléculas. Pronto se verá tan sacudido por el movimiento browniano que no puede decir si viene o va, y mucho menos si las moléculas vienen o van, por lo que no sirve.

46-4 Irreversibilidad

¿Son reversibles todas las leyes de la física? ¡Evidentemente no! ¡Traten solamente de desbatir un huevo! Pasen una película hacia atrás, y solamente se tarda unos minutos para que todo el mundo comience a reír. La característica más natural de todos los fenómenos es su obvia irreversibilidad.

¿De dónde proviene la irreversibilidad? No de las leyes de Newton. Si sostenemos que el comportamiento de todo se tiene que entender en último término mediante las leyes de la física, y si también resulta que todas las ecuaciones tienen la fantástica propiedad de que si ponemos $t = -t$ tenemos otra solución, entonces todo fenómeno es reversible. ¿Cómo es entonces que en la naturaleza en gran escala, las cosas no son reversibles? Obviamente debe haber alguna ley, alguna ecuación fundamental aunque oscura, quizás en electricidad, quizás en la física del neutrino, en la cual sí *importa* de qué modo marcha el tiempo.

Discutamos esta pregunta ahora. Ya conocemos una de estas leyes, la que dice que la entropía siempre está aumentando. Si tenemos una cosa caliente y otra fría, el calor pasa de la caliente a la fría. Por esto la ley de la entropía es una de tales leyes. Pero esperamos entender la ley de la entropía desde el punto de vista de la mecánica. De hecho, acabamos de tener éxito al obtener todas las consecuencias del razonamiento de que el calor no puede fluir hacia atrás por sí mismo exclusivamente por razones mecánicas, y con ello obtuvimos una comprensión de la Segunda ley. Aparentemente podemos obtener irreversibilidad de ecuaciones reversibles. Pero *¿fue solamente* un razonamiento mecánico el que usamos? Considerémoslo con más cuidado.

Ya que nuestra pregunta tiene que ver con la entropía, el problema es tratar de encontrar una descripción microscópica de la entropía. Si decimos que tenemos una cierta cantidad de energía en algo, como en un gas, entonces podemos obtener una imagen microscópica de ello y decir que cada átomo tiene una cierta energía. Todas estas energías sumadas nos dan la energía total. Del mismo modo, quizás, cada átomo tiene una cierta entropía. Si lo sumamos todo, tendremos la entropía total. La cosa no es tan fácil, pero veamos lo que sucede.

Como ejemplo, calculemos la diferencia de entropía entre un gas a una cierta temperatura a un volumen, y un gas a la misma temperatura a otro volumen. Recordemos que en el capítulo 44 teníamos, para la variación de entropía:

$$\Delta S = \int \frac{dQ}{T}.$$

En el caso presente, la energía del gas es la misma antes y después de la expansión, ya que la temperatura no varía. De manera que tenemos que añadir calor suficiente para igualar el trabajo efectuado por el gas, o sea, para un cambio pequeño en volumen, tenemos:

$$dQ = P\,dV.$$

Poniendo esto en dQ, obtenemos:

$$\begin{aligned}\Delta S &= \int_{V_1}^{V_2} P\,\frac{dV}{T} = \int_{V_1}^{V_2} \frac{NkT}{V}\,\frac{dV}{T}\\ &= Nk\ln\frac{V_2}{V_1},\end{aligned}$$

como obtuvimos en el capítulo 44. Por ejemplo, si expandimos el volumen en un factor 2, la variación de entropía es $Nk\ln 2$.

Consideremos ahora otro ejemplo interesante. Supongan que tenemos una caja con una barrera en el medio. En un lado hay neón (moléculas “negras”) y en el otro argón (moléculas “blancas”). Quitamos ahora la barrera y permitimos que se mezclen. ¿Cuánto ha variado la entropía? Es posible imaginar que en lugar de la barrera tenemos un pistón con agujeros que permiten que pasen las blancas pero no las negras, y otro pistón que hace lo contrario. Si colocamos cada pistón en un extremo, vemos que para cada gas el problema es parecido al que acabamos de resolver. Así, pues, obtenemos una variación de entropía de $Nk\ln 2$, lo que significa que la entropía ha aumentado en $k\ln 2$ por molécula. El 2 tiene que ver con el espacio adicional que posee la molécula, lo cual es bastante extraño. No es una propiedad de la misma molécula, sino de *cuánto espacio* tiene la molécula para moverse. Es una situación extraña donde la entropía aumenta, pero ¡donde todo tiene la misma temperatura y la misma energía! La única cosa que ha variado es que las moléculas están distribuidas de otra manera.

Sabemos bien que si solamente quitamos la barrera, todo estará mezclado después de un tiempo largo, debido a las colisiones, a las vibraciones, a los golpes, etc. De vez en cuando una molécula blanca va hacia una negra y una negra va hacia una blanca y puede ser que se crucen. Las blancas gradualmente se abren paso, accidentalmente, en el espacio de las negras, y las negras se abren paso, accidentalmente, en el espacio de las blancas. Si esperamos lo suficiente obtendremos una mezcla. Claramente, éste es un proceso irreversible en el mundo real y debería implicar un aumento de entropía.

Aquí tenemos un ejemplo simple de un proceso irreversible que está completamente formado de eventos reversibles. Cada vez que hay una colisión entre dos moléculas cualesquiera, salen en ciertas direcciones. Si tomamos una película de una colisión hacia atrás, no habría nada erróneo con la película. En realidad, una clase de colisión es tan probable como otra. Por lo que la mezcla es completamente reversible y no obstante es irreversible. Todo el mundo sabe que si comenzásemos con blancas y negras, separadas, obtendríamos una mezcla en unos pocos minutos. Si nos sentáramos y siguiéramos contemplándolas por unos cuantos minutos más, no se separarían de nuevo, sino que permanecerían mezcladas. Así nos encontramos con una irreversibilidad basada en situaciones reversibles. Pero también vemos ahora la *razón.* Empezamos con una distribución que está, en algún sentido, *ordenado.* Debido al caos de las colisiones, se desordena. *La fuente de la irreversibilidad es el cambio de una distribución ordenada a una desordenada.*

Es verdad que si tomásemos una película de esto y la proyectásemos hacia atrás veríamos que gradualmente se ordena. Alguien diría: “¡Esto es contra las leyes de la física!” Pasamos la película de nuevo y miramos cada colisión. Cada una sería perfecta y cada una estaría obedeciendo las leyes de la física. La razón es, naturalmente, que las velocidades de cada molécula son exactamente correctas, por lo que si seguimos los caminos hacia atrás vuelven a su condición inicial. Pero esto es una circunstancia muy poco probable. Si empezamos con el gas sin distribución especial, sólo blancas y negras, nunca volverá atrás.

46-5 Orden y entropía

Ahora tenemos, pues, que hablar de lo que entendemos por orden y desorden. No es cuestión de un orden agradable o de un desorden desagradable. Lo que es diferente en nuestros casos mezclados y sin mezclar es lo siguiente. Supongan que

dividimos el espacio en pequeños elementos de volumen. Si tenemos moléculas blancas y negras, ¿de cuántas maneras podríamos distribuirlas entre los elementos de volumen de modo que las blancas estuviesen en un lado y las negras en el otro? Por otro lado, ¿de cuántas maneras podríamos distribuirlas sin ninguna restricción de dónde va cada una? Claramente, hay muchas más maneras de colocarlas en el último caso. Medimos "desorden" por el número de maneras en que podemos disponer las cosas internamente, de modo que parezca lo mismo desde el exterior. *La entropía es el logaritmo de ese número de maneras.* El número de maneras en el caso de los gases separados es menor, por lo que la entropía es menor, o el "desorden" es menor.

Así, con la precedente definición técnica de desorden podemos entender la proposición. En primer lugar, la entropía mide el desorden. En segundo lugar, el universo siempre va de "orden" a "desorden", por lo que la entropía siempre aumenta. No es un orden en el sentido de que el arreglo nos agrada, sino en el sentido de que el número de maneras diferentes en que lo podemos lograr, conservando el mismo aspecto desde el exterior, es relativamente restringido. En el caso en que invertíamos la película del gas mezclándose, no había tanto desorden como pensábamos. ¡Cada uno de los átomos tenía exactamente la velocidad y dirección correcta para salir airoso! La entropía no era alta después de todo, aunque parecía serlo.

¿Qué hay de la reversibilidad de las otras leyes físicas? Al hablar del campo eléctrico producido por una carga acelerada, se dijo que debíamos tomar este campo retardado. A un tiempo t y a una distancia r de la carga, tomamos el campo debido a la aceleración al tiempo $t - r/c$, y no a $t + r/c$. Así parece, a primera vista, como que la ley de la electricidad no es reversible. Muy extrañamente, sin embargo, las leyes que usamos provienen de un conjunto de ecuaciones llamadas ecuaciones de Maxwell que son, en realidad, reversibles. Además, es posible argumentar que si solamente tuviéramos que usar el campo adelantado, el debido al estado de cosas a $t + r/c$, y hacerlo en forma absolutamente compatible en un espacio completamente cerrado, ¡todo sucede exactamente del mismo modo que si usamos campos retardados! Esta aparente irreversibilidad en electricidad, al menos en un recinto, no es pues una irreversibilidad de ninguna especie. Ya tenemos alguna indicación de esto, porque sabemos que cuando tenemos una carga oscilante que genera campos que son rechazados por las paredes de un recinto, llegamos finalmente a un equilibrio en el que no hay ninguna unilateralidad. La aproximación del campo retardado es solamente una conveniencia en el método de la solución.

En cuanto a lo que sabemos, todas las leyes fundamentales de la física, como las ecuaciones de Newton, son reversibles. Entonces ¿de dónde sale la irreversibilidad? Viene del orden volviéndose desorden, pero no entenderemos esto hasta que sepamos el origen del orden. ¿Por qué las situaciones en que nos encontramos diariamente están siempre fuera del equilibrio? Una explicación posible es la siguiente. Examinemos de nuevo nuestra caja de moléculas blancas y negras mezcladas. Bien, si esperamos un tiempo suficiente, es posible que por puro accidente, muy improbable pero posible, la distribución de moléculas llegue a ser principalmente de blancas en un lado y principalmente de negras en el otro. Después de esto, a medida que pasa el tiempo y los accidentes continúan, se vuelven a mezclar de nuevo.

Entonces, una explicación posible del alto grado de orden en el mundo de hoy es que es solamente una cuestión de suerte. Quizá sucedió que nuestro mundo tuvo en el pasado una fluctuación de alguna clase en la que las cosas se separaron de algún modo y ahora están volviendo a unirse de nuevo. Esta clase de teoría no es asimétrica, porque podemos preguntar qué aspecto tiene el gas separado o un poquito en el futuro o un poquito en el pasado. En uno u otro caso, vemos una mancha gris en la interfase debido a que las moléculas se están mezclando de nuevo. El gas se mezcla, cualquiera que sea el sentido en que transcurre el tiempo. Esta teoría diría, pues, que la irreversibilidad es solamente uno de los accidentes de la vida.

Queremos argumentar que éste no es el caso. Supongan que no miramos toda la caja a la vez, sino únicamente una parte de ella. Supongan, entonces, que en un cierto momento descubrimos una cierta cantidad de orden. En este trozo pequeño, las blancas y las negras están separadas. ¿Qué deduciríamos de la situación en los lugares donde aún no hemos mirado? Si creemos realmente que el orden proviene del desorden completo debido a una fluctuación, debemos seguramente tomar la fluctuación, más probable que podría producirla, y la condición más probable ¡*no* es que el resto también se haya desenredado! Por consiguiente, a partir de la hipótesis de que el mundo es una fluctuación, todas las predicciones son que, si miramos a una parte del mundo que nunca hemos visto antes, la encontraremos mezclada y no como el trozo que acabamos de considerar. Si nuestro orden se debiese a una fluctuación, no esperaríamos orden en ninguna parte, sino solamente donde lo hemos notado.

Ahora suponemos que la separación es porque el pasado del universo estaba realmente ordenado. No es debido a una fluctuación, sino que todo solía ser blanco y negro. Esta teoría predice que habrá orden en otros lugares —el orden no se debe a una fluctuación, sino a un ordenamiento más alto al comienzo de los tiempos—. Entonces esperaríamos encontrar orden en lugares donde aún no hemos mirado.

Los astrónomos, por ejemplo, han mirado solamente algunas estrellas. Todos los días enfocan sus telescopios a otras estrellas y las nuevas estrellas están haciendo lo mismo que las otras. Por lo tanto, nosotros concluimos que el universo *no* es una fluctuación y que el orden es un recuerdo de las condiciones cuando todo empezó. Esto no es decir que entendemos su lógica. Por alguna razón, el universo tuvo en un tiempo una entropía muy baja para su contenido de energía y desde entonces la entropía ha aumentado. Y así, éste es el camino hacia el futuro. Éste es el origen de toda irreversibilidad, esto es lo que hace el proceso de crecimiento y decadencia, esto lo que nos hace recordar el pasado y no el futuro, recordar las cosas que están más cercanas a aquel momento en la historia del universo cuando el orden era mayor que ahora, y el por qué no podemos recordar cosas donde el desorden es mayor que ahora, a las cuales llamamos futuro. Por ello, como comentábamos en un capítulo anterior, el universo entero está en un vaso de vino, si lo miramos suficientemente cerca. En este caso el vaso de vino es complejo porque hay agua, vidrio, luz y todo lo demás.

Otra de las delicias de la física es que aun las cosas simples e idealizadas, como la rueda dentada y trinquete, trabajan solamente porque son parte del universo. La rueda dentada y trinquete funcionan únicamente en una dirección porque tienen un contacto fundamental con el resto del universo. Si estuviera en una caja aislada durante un tiempo suficiente, la rueda no tendría más probabilidad de girar en un sentido que en el otro. Pero, porque quitamos las sombras y dejamos la luz, porque nos enfriamos en la Tierra y obtenemos calor del Sol, las ruedas dentadas y trinquetes que construimos pueden girar en un sentido. Este sentido único está interrelacionado con el hecho de que la rueda dentada es parte del universo. Es parte del universo no solamente en el sentido de que obedece las leyes físicas del universo, sino que su comportamiento de sentido único está ligado al comportamiento de sentido único del universo entero. Esto no se podrá entender completamente hasta que el misterio de los comienzos de la historia del universo se reduzca aún más de la especulación a la comprensión científica.

47

Sonido. La ecuación de onda

47-1 Ondas

En este capítulo estudiaremos el fenómeno de las *ondas*. Es un fenómeno que aparece en muchos contextos en toda la física, por lo que nuestra atención se debe concentrar en él, no sólo por el ejemplo particular considerado aquí, el sonido, sino también por el amplio campo de aplicación de estas ideas en todas las ramas de la física.

Al estudiar el oscilador armónico señalamos que no sólo hay ejemplos mecánicos de sistemas oscilantes, sino también eléctricos. Las ondas están relacionadas con sistemas oscilantes, sólo que las oscilaciones ondulatorias, además de aparecer como oscilaciones temporales en un punto, se propagan en el espacio.

En realidad, ya hemos estudiado ondas. Cuando estudiábamos la luz, al enterarnos de las propiedades pertinentes de las ondas, prestamos atención en particular a la interferencia espacial de ondas provenientes de varias fuentes colocadas en diferentes puntos, todas de la misma frecuencia. Hay dos fenómenos ondulatorios importantes que aún no hemos estudiado y que ocurren tanto en la luz, es decir, ondas electromagnéticas, como en cualquier otro tipo de ondas. El primero es el fenómeno de *interferencia temporal* en vez de interferencia espacial. Si tenemos dos fuentes de sonido con frecuencias ligeramente diferentes y escuchamos las dos al mismo tiempo, algunas veces las ondas vienen con las crestas juntas y otras con la cresta y el valle juntos (véase la figura 47-1). El aumento y disminución del sonido que resulta es el fenómeno de *pulsación*, o sea de interferencia en el tiempo. El segundo fenómeno es el concerniente a los diagramas de ondas que resultan cuando las ondas están confinadas en un volumen determinado y se reflejan de un lado a otro en las paredes.

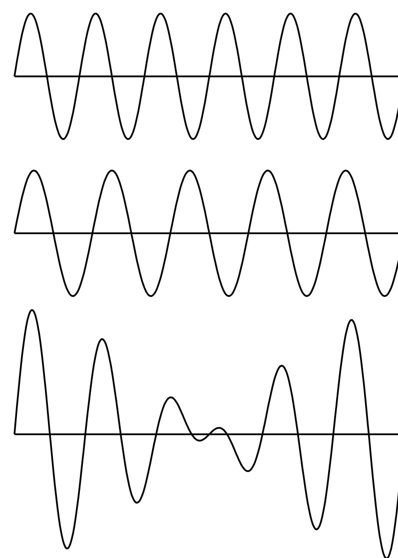


Figura 47-1. Interferencia temporal de dos ondas sonoras de frecuencia ligeramente diferente, dando lugar a pulsaciones.

Desde luego que podríamos haber discutido estos fenómenos en el caso de las ondas electromagnéticas. La razón de no haberlo hecho es que usando un solo ejemplo no hubiéramos provocado la sensación de estar verdaderamente aprendiendo muchos temas diferentes al mismo tiempo. Para hacer resaltar la aplicabilidad general de las ondas más allá de la electrodinámica, consideramos aquí un ejemplo diferente, en particular ondas sonoras.

Otros ejemplos de ondas son las ondas de agua consistentes en los largos abultamientos que vemos acercarse a la costa, o las ondas de agua más pequeñas consistentes en ondulaciones por tensión superficial. Como otro ejemplo están las dos clases de ondas elásticas en sólidos, una onda de compresión (o longitudinal) en la que las partículas del sólido oscilan en la dirección de propagación de la onda (las ondas sonoras en un gas son de este tipo), y una onda transversal en la que las partículas del sólido oscilan perpendicularmente a la dirección de propagación. Las ondas de los terremotos contienen las dos clases de ondas elásticas, generadas por un movimiento en algún punto de la corteza terrestre.

Un ejemplo de ondas se encuentra en la física moderna. Éstas son ondas que dan la amplitud de probabilidad de encontrar una partícula en un lugar determinado —las "ondas de materia" que ya hemos discutido—. Su frecuencia es proporcional a la energía y su número de onda es proporcional al momentum. Son las ondas de la mecánica cuántica.

En este capítulo consideraremos únicamente ondas cuya velocidad es independiente de la longitud de onda. Éste es el caso, por ejemplo, de la luz en el vacío. La velocidad es entonces la misma para las ondas de radio, para la luz azul, para la luz verde o para cualquier otra longitud de onda. A causa de este comportamiento, cuando comenzamos a describir los fenómenos ondulatorios no

nos dimos cuenta al principio de que teníamos propagación de ondas. En vez de ello dijimos que si una carga se mueve en un lugar, el campo eléctrico a una distancia x era proporcional a la aceleración, pero no a un tiempo t sino a un tiempo anterior $t - x/c$. En consecuencia, si fuéramos a dibujar el campo eléctrico en el espacio en cierto instante, como en la figura 47-2, el campo eléctrico a un tiempo t posterior se habrá desplazado la distancia ct, como se indica en la figura. Matemáticamente podemos decir que, en el ejemplo unidimensional de que hablamos, el campo eléctrico es función de $x - ct$. Vemos que $t = 0$ es cierta función de x. Si consideramos un tiempo posterior, sólo necesitamos aumentar x un poco para obtener el mismo valor del campo eléctrico. Por ejemplo, si el campo máximo está en $x = 3$ con tiempo cero, para encontrar la nueva posición del campo máximo al tiempo t necesitamos:

$$x - ct = 3 \qquad \text{o} \qquad x = 3 + ct.$$

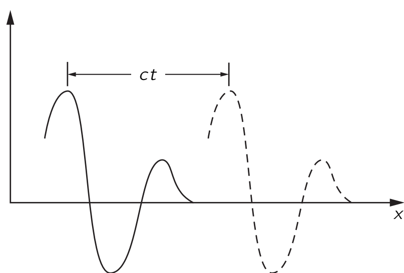


Figura 47-2. La curva continua muestra el aspecto que podría tener el campo eléctrico en cierto instante y la curva discontinua muestra cuál es el campo eléctrico un tiempo t más tarde.

Vemos que este tipo de función representa la propagación de una onda.

Esa función $f(x - ct)$, representa, pues, una onda. Podemos resumir esta descripción de una onda diciendo simplemente que

$$f(x - ct) = f(x + \Delta x - c(t + \Delta t)),$$

cuando $\Delta x = c\ \Delta t$. Hay, por supuesto, otra posibilidad: que en vez de una fuente a la izquierda, como se indica en la figura 47-2, tengamos una fuente a la derecha, de modo que la onda se propague hacia las x negativas. Entonces, la onda estaría descrita por $g(x + ct)$.

Hay otra posibilidad: que exista más de una onda en el espacio al mismo tiempo; el campo eléctrico es entonces la suma de los dos campos, cada uno propagándose independientemente. Se puede describir este comportamiento de los campos eléctricos diciendo que si $f_1(x - ct)$ es una onda y si $f_2(x - ct)$ es otra onda, su suma también es una onda. Esto se denomina principio de superposición. El mismo principio es válido para el sonido.

Estamos familiarizados con el hecho de que si se produce un sonido, escuchamos con completa fidelidad la misma secuencia de sonidos que se generó. Si tuviéramos altas frecuencias viajando más rápido que las bajas, un ruido breve y definido se oiría como una sucesión de sonidos musicales. Análogamente, si la luz roja viajara más rápido que la azul, un destello de luz blanca se vería primero rojo, luego blanco y finalmente azul. La experiencia diaria nos dice que no es así. Tanto el sonido como la luz viajan en el aire con una velocidad que es casi totalmente independiente de la frecuencia. En el capítulo 48 consideraremos ejemplos de propagación de ondas para los cuales no se verifica esta independencia.

En el caso de la luz (ondas electromagnéticas) dimos una regla para determinar un punto del campo eléctrico que resulta de la aceleración de una carga. Sería de esperar ahora que lo que debemos hacer es dar una regla para determinar alguna cualidad del aire, la presión digamos, a cierta distancia de una fuente en términos del movimiento de la fuente, retardada en el tiempo de viaje del sonido. En el caso de la luz, este procedimiento era aceptable porque todo lo que sabíamos era que una carga en un lugar ejerce una fuerza sobre otra carga en otro lugar. Los detalles de la propagación de un lugar a otro no eran esenciales en absoluto. Por el contrario, en el caso del sonido sabemos que se propaga por el aire entre la fuente y el que escucha, y es totalmente natural que se pregunte cuál es la presión del aire en cualquier momento determinado. Además, querríamos saber exactamente cómo se mueve el aire. En el caso de la electricidad podíamos aceptar una regla, puesto que podíamos decir que todavía no conocíamos las leyes de la electricidad, pero no podemos decir lo mismo respecto al sonido. No quedaríamos satisfechos con una regla que estableciese cómo se mueve por el aire la presión del sonido, porque se debería comprender el proceso como consecuencia de las leyes de la mecánica. En suma, el sonido es una rama de la mecánica y por lo tanto hay que comprenderlo con las leyes de Newton. La propagación del sonido de un lugar a otro es sencillamente una consecuencia de la mecánica y de las propiedades de los gases, si se propaga en un gas, o de las propiedades de

líquidos o sólidos, si se propaga en esos medios. Más adelante deduciremos las propiedades de la luz y su propagación ondulatoria en forma análoga a partir de las leyes de la electrodinámica.

47-2 Propagación del sonido

Daremos una derivación de las propiedades de la propagación del sonido *entre* la fuente y el receptor como una consecuencia de las leyes de Newton y nosotros no consideraremos la interacción con la fuente y el receptor. Ordinariamente recalcamos un resultado en lugar de una derivación particular de él. En este capítulo adoptamos el punto de vista opuesto. Aquí lo importante es, en cierto sentido, la derivación en sí. Este problema de explicar nuevos fenómenos en función de los viejos, cuando conocemos las leyes de los viejos, es posiblemente el mayor arte de la física matemática. El físico matemático tiene dos problemas: uno es encontrar soluciones dadas las ecuaciones y otro es encontrar las ecuaciones que describen un fenómeno nuevo. La derivación que daremos aquí es un ejemplo del segundo tipo de problema.

Tomaremos aquí el ejemplo más simple: la propagación del sonido en una dimensión. Para llevar a cabo esa derivación es necesario tener primero algún conocimiento de lo que está pasando. Fundamentalmente, lo que interviene es que si se mueve un objeto en un punto en el aire, observamos que hay una perturbación que viaja por el aire. Si preguntamos qué clase de perturbación, diríamos que es de esperar que el movimiento del objeto produzca un cambio de presión. Desde luego, si el objeto se mueve suavemente, el aire fluye a su alrededor simplemente, pero lo que nos interesa es un movimiento rápido, de modo que no hay tiempo suficiente para ese flujo. Por lo tanto, el aire se comprime con el movimiento y se produce un cambio de presión que empuja más aire. A su vez, este aire se comprime, lo cual vuelve a dar una presión adicional, y se propaga una onda.

Ahora queremos formular tal proceso. Tenemos que decidir qué variables necesitamos. En nuestro problema necesitaríamos saber cuánto se ha movido el aire, por lo que el *desplazamiento* del aire en la onda sonora es ciertamente una variable relevante. Además, querríamos describir cómo cambia la *densidad* del aire a medida que se desplaza. La *presión* del aire también cambia, así que ésta es otra variable de interés. Naturalmente, el aire también tiene una *velocidad*, por lo que tendremos que describir la velocidad de las partículas del aire. Las partículas del aire también tienen *aceleraciones* —pero al dar la lista de todas estas variables, pronto nos damos cuenta de que se conocería la velocidad y la aceleración si supiéramos cómo varía en el tiempo el *desplazamiento* del aire.

Como dijimos, consideraremos la onda en una dimensión. Lo podemos hacer si estamos suficientemente lejos de la fuente como para que lo que llamamos *frentes frente de* sean casi planos. De esta manera simplificamos nuestro razonamiento tomando el ejemplo menos complicado. Entonces podremos decir que el desplazamiento, χ, sólo depende de x y de t y no de y o z. En consecuencia, la descripción del aire está dada por $\chi(x, t)$.

¿Es completa esta descripción? Parecería que está muy lejos de serlo, puesto que no conocemos ningún detalle de cómo se mueven las moléculas del aire. Se están moviendo en todas direcciones y ciertamente este estado de cosas no está descrito con esta función $\chi(x, t)$. Desde el punto de vista de la teoría cinética, si tenemos una densidad más alta de moléculas en un lugar y una más baja en la región adyacente, las moléculas se mudarán de la región de alta densidad a la de baja, de modo que la diferencia se compense. Aparentemente, no obtendríamos una oscilación y no habría sonido. Lo que se necesita para obtener la onda sonora es esta situación: cuando las moléculas se precipitan fuera de la región de densidad y presión más altas, entregan momentum a las moléculas de la región adyacente de densidad más baja. Para que se genere sonido, las regiones en las que varían la densidad y la presión deben ser mucho mayores que la distancia recorrida por las moléculas antes de chocar con otras moléculas. Esta distancia es el camino libre medio y la distancia entre las crestas y los valles de presión debe

ser mucho mayor que él. De otro modo las moléculas se moverían libremente de la cresta al valle, alisando inmediatamente la onda.

Está claro que vamos a describir el comportamiento del gas en una escala grande respecto al camino libre medio, por lo que las propiedades del gas no estarán descritas en términos de moléculas individuales. El desplazamiento, por ejemplo, será el desplazamiento del centro de masa de un pequeño elemento del gas y la presión o la densidad serán la presión o la densidad en esta región. Llamaremos P a la presión y ρ a la densidad, y ambas serán funciones de x y de t. No debemos olvidar que esta descripción es una aproximación válida únicamente cuando estas propiedades del gas no varían demasiado rápidamente con la distancia.

47-3 La ecuación de onda

La física del fenómeno de las ondas sonoras comprende entonces tres características:

I. El gas se mueve y varía la densidad.

II. La variación de densidad corresponde a una variación de presión.

III. Las desigualdades de presión generan el movimiento del gas.

Consideremos primero el punto II. Para un gas, un líquido o un sólido, la presión es cierta función de la densidad. Antes de que llegue la onda sonora tenemos equilibrio a una presión P_0 y una densidad correspondiente ρ_0. Una presión P del medio está ligada con la densidad por una relación característica $P = f(\rho)$ y, en particular, la presión P_0 de equilibrio está dada por $P_0 = f(\rho_0)$. En el sonido, las variaciones de presión respecto al valor de equilibrio son extremadamente pequeñas. Una unidad conveniente para medir presiones es el *bar*, siendo 1 bar igual a 10^5 n/m^2. La presión de 1 atmósfera patrón es muy cercana a 1 bar, 1 atm = 1.013 bar. Para el sonido usamos una escala logarítmica de intensidades ya que la sensibilidad del oído es aproximadamente logarítmica. Esta escala es la escala en decibelios, en la cual el nivel de presión acústica para la amplitud de presión P se define por:

$$I \text{ (nivel de presión acústica)} = 20 \log_{10}(P/P_{\text{ref}}) \text{ en dB,} \tag{47.1}$$

donde la presión de referencia $P_{\text{ref}} = 2 \times 10^{-10}$ bar. Una amplitud de presión $P = 10^3 P_{\text{ref}} = 2 \times 10^{-7}$ bar* corresponde a un sonido moderadamente intenso de 60 decibelios. Vemos que en el sonido las variaciones de presión son extremadamente pequeñas respecto a la presión de equilibrio, o media, de 1 atm. Correspondientemente, los desplazamientos y las variaciones de densidad son extremadamente pequeños. En las explosiones no tenemos esas pequeñas variaciones, la presión adicional producida puede ser mayor de 1 atm. Estas grandes variaciones de presión dan lugar a efectos nuevos que consideraremos más adelante. En el sonido pocas veces consideramos niveles de intensidad acústica por encima de 100 dB; 120 dB es un nivel doloroso para el oído. Para el sonido, en consecuencia, si escribimos

$$P = P_0 + P_e, \qquad \rho = \rho_0 + \rho_e, \tag{47.2}$$

siempre tendremos que la variación de presión P_e es muy pequeña frente a P_0 y que la variación ρ_e de densidad es muy pequeña respecto a ρ_0. Luego,

$$P_0 + P_e = f(\rho_0 + \rho_e) = f(\rho_0) + \rho_e f'(\rho_0), \tag{47.3}$$

donde $P_0 = f(\rho_0)$ y $f'(\rho_0)$ es el valor de la derivada de $f(\rho)$ para $\rho = \rho_0$. Podemos detenernos en el segundo paso de esta igualdad únicamente porque ρ_e es muy

* Con esta elección de P_{ref}, la P no es la presión de pico en la onda sonora sino la presión "media cuadrática", que es $1/(2)^{1/2}$ por la presión de pico.

pequeña. De este modo encontramos que la presión P_e en exceso es proporcional a la densidad ρ_e en exceso, pudiendo llamar κ al factor de proporcionalidad,

$$P_e = \kappa\rho_e, \qquad \text{donde } \kappa = f'(\rho_0) = (dP/d\rho)_0. \tag{47.4}$$

Esta relación simplísima es la que necesitábamos para II.

Ahora consideremos I. Supondremos que x es la posición de una porción de aire no perturbada por la onda sonora y que $\chi(x, t)$ es el desplazamiento al tiempo t debido al sonido, por lo que la nueva posición es $x + \chi(x, t)$, como en la figura 47-3. Ahora bien, la posición no perturbada de una porción vecina de aire es $x + \Delta x$, y la nueva posición es $x + \Delta x + \chi(x + \Delta x, t)$. Podemos ahora hallar las variaciones de densidad en la siguiente forma. Como nos estamos limitando a ondas planas, podemos tomar un área unitaria perpendicular a la dirección x, que es la dirección de propagación de la onda sonora. La cantidad de aire por unidad de área en Δx es entonces $\rho_0\Delta x$, donde ρ_0 es la densidad del aire no perturbada o de equilibrio. Este aire, cuando la onda sonora lo desplaza, queda entre $x + \chi(x, t)$ y $x + \Delta x + \chi(x + \Delta x, t)$, de modo que en este intervalo tenemos la misma materia que había en Δx cuando no había perturbación. Si ρ es la nueva densidad, se tiene:

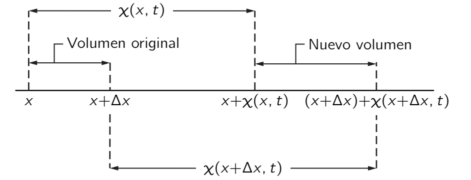


Figura 47-3. El desplazamiento del aire en x es $\chi(x, t)$, y en $x + \Delta x$ es $\chi(x + \Delta x, t)$. El volumen original de aire para un área unitaria de la onda plana es Δx; el nuevo volumen es $\Delta x + \chi(x + \Delta x, t) - \chi(x, t)$.

$$\rho_0\,\Delta x = \rho[x + \Delta x + \chi(x + \Delta x, t) - x - \chi(x, t)]. \tag{47.5}$$

Como Δx es pequeño, podemos escribir $\chi(x + \Delta x, t) - \chi(x, t) = (\partial\chi/\partial x)\Delta x$. Esta derivada es una derivada parcial, puesto que χ depende tanto del tiempo como de x. Nuestra ecuación es entonces:

$$\rho_0\,\Delta x = \rho\left(\frac{\partial\chi}{\partial x}\,\Delta x + \Delta x\right) \tag{47.6}$$

o sea

$$\rho_0 = (\rho_0 + \rho_e)\frac{\partial\chi}{\partial x} + \rho_0 + \rho_e. \tag{47.7}$$

Ahora bien, en las ondas sonoras todas las variaciones son pequeñas, de modo que ρ_e es pequeña, χ es pequeña y $\partial\chi/\partial x$ también es pequeña. En consecuencia, en la relación que acabamos de encontrar,

$$\rho_e = -\rho_0\,\frac{\partial\chi}{\partial x} - \rho_e\,\frac{\partial\chi}{\partial x}, \tag{47.8}$$

podemos despreciar $\rho_e\partial\chi/\partial x$ frente a $\rho_0\partial\chi/\partial x$. Obtenemos así la relación que necesitábamos para I:

$$\rho_e = -\rho_0\,\frac{\partial\chi}{\partial x}. \qquad \text{(I)} \tag{47.9}$$

Esta ecuación es la que sería de esperar físicamente. Si los desplazamientos varían con x, habrá variaciones de densidad. También el signo es correcto, si el desplazamiento χ aumenta con x, de modo que el aire se estira, la densidad tiene que disminuir.

Ahora necesitamos la tercera ecuación, que es la del movimiento producido por la presión. Si conocemos la relación entre la fuerza y la presión, podemos obtener la ecuación de movimiento. Si tomamos una rebanada de aire de espesor Δx y de área unitaria perpendicular a x, la masa de aire de esta rebanada es $\rho_0\Delta x$ y tiene una aceleración $\partial^2\chi/\partial t^2$, por lo que la masa por la aceleración de esta rebanada de materia es $\rho_0\Delta x(\partial^2\chi/\partial t^2)$. (Para Δx pequeño no importa que evaluemos la aceleración $\partial^2\chi/\partial t^2$ en un borde de la rebanada o en una posición intermedia.) Si ahora encontramos la fuerza que actúa sobre esta materia por unidad de área perpendicular a x, será igual a $\rho_0\Delta x(\partial^2\chi/\partial t^2)$. En x tenemos la fuerza en la dirección $+x$, de módulo $P(x, t)$ por unidad de área, y en $x + \Delta x$ tenemos la fuerza en dirección opuesta, de módulo $P(x + \Delta x, t)$ por unidad de área (figura 47-4)

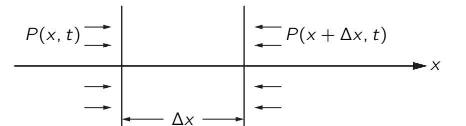


Figura 47-4. La fuerza resultante en la dirección x positiva producida por la presión que actúa sobre un área unitaria perpendicular a x, es $-(\partial P/\partial x)\,\Delta x$.

$$P(x, t) - P(x + \Delta x, t) = -\frac{\partial P}{\partial x}\,\Delta x = -\frac{\partial P_e}{\partial x}\,\Delta x, \tag{47.10}$$

ya que Δx es pequeño y que la única parte de P que varía es la presión en exceso P_e. Ahora tenemos III:

$$\rho_0 \frac{\partial^2 \chi}{\partial t^2} = -\frac{\partial P_e}{\partial x}, \qquad \text{(III)} \tag{47.11}$$

y así tenemos ecuaciones suficientes para interconectar las cosas y reducir a una variable, a χ digamos. Podemos eliminar P_e de III usando II, con lo que obtenemos:

$$\rho_0 \frac{\partial^2 \chi}{\partial t^2} = -\kappa \frac{\partial \rho_e}{\partial x}, \tag{47.12}$$

y luego podemos usar I para eliminar ρ_e. De este modo encontramos que ρ_0 se simplifica y nos queda:

$$\frac{\partial^2 \chi}{\partial t^2} = \kappa \frac{\partial^2 \chi}{\partial x^2}. \tag{47.13}$$

Llamaremos $c_s^2 = \kappa$, por lo que podemos escribir:

$$\frac{\partial^2 \chi}{\partial x^2} = \frac{1}{c_s^2} \frac{\partial^2 \chi}{\partial t^2}. \tag{47.14}$$

Ésta es la ecuación de onda que describe el comportamiento del sonido en la materia.

47-4 Soluciones de la ecuación de onda

Ahora podemos ver si esta ecuación describe realmente las propiedades esenciales de las ondas sonoras en la materia. Queremos deducir que un pulso sonoro, o perturbación, se moverá con velocidad constante. Queremos verificar que dos pulsos diferentes se pueden mover uno a través del otro —el principio de superposición—. También queremos verificar que el sonido puede ir tanto a la derecha como a la izquierda. Todas estas propiedades deberían estar contenidas en esta única ecuación.

Hemos hecho notar que cualquier perturbación tipo onda plana que se mueve con velocidad constante v tiene forma $f(x - vt)$. Ahora tenemos que ver si $\chi(x, t) = f(x - vt)$ es solución de la ecuación de onda. Al calcular $\partial\chi/\partial x$, obtenemos la derivada de la función: $\partial\chi/\partial x = f'(x - vt)$. Derivando una vez más, encontramos:

$$\frac{\partial^2 \chi}{\partial x^2} = f''(x - vt). \tag{47.15}$$

La derivación de esta función respecto a t da $-v$ por la derivada de la función, o sea $\partial\chi/\partial t = -vf'(x - vt)$, y la segunda derivada respecto al tiempo es

$$\frac{\partial^2 \chi}{\partial t^2} = v^2 f''(x - vt). \tag{47.16}$$

Es evidente que $f(x - vt)$ satisfará la ecuación de onda siempre que la velocidad de onda v sea igual a c_s.

En consecuencia, partiendo de las *leyes de la mecánica* encontramos que cualquier perturbación sonora se propaga con velocidad c_s y además:

$$c_s = \kappa^{1/2} = (dP/d\rho)_0^{1/2},$$

y así *hemos relacionado la velocidad de onda con una propiedad del medio.*

Si consideramos una onda que viaja en dirección opuesta, de modo que $\chi(x, t) = g(x + vt)$, es fácil ver que esa perturbación también satisface la ecuación de onda. La única diferencia entre una onda de este tipo y una que viaja de izquierda a derecha es el signo de v, pero sea que tengamos $x + vt$ o $x - vt$ como variable en la función, el signo de $\partial^2\chi/\partial t^2$ no se altera, ya que en ésta sólo

interviene v^2. Se sigue que tenemos soluciones para ondas que se propagan en una u otra dirección con velocidad c_s.

Una cuestión extremadamente importante es la de la superposición. Supongan que se ha encontrado una solución de la ecuación de onda, χ_1 digamos. Esto significa que la segunda derivada de χ_1 respecto a x es $1/c_s^2$ por la segunda derivada de χ_1 respecto a t. Ahora bien, cualquier otra solución χ_2 tiene la misma propiedad. Si superponemos estas dos soluciones tenemos:

$$\chi(x,t) = \chi_1(x,t) + \chi_2(x,t), \tag{47.17}$$

y deseamos verificar que $\chi(x,t)$ también es una onda, es decir, que satisface χ la ecuación de onda. Podemos demostrar fácilmente este resultado, puesto que tenemos:

$$\frac{\partial^2\chi}{\partial x^2} = \frac{\partial^2\chi_1}{\partial x^2} + \frac{\partial^2\chi_2}{\partial x^2} \tag{47.18}$$

y, además,

$$\frac{\partial^2\chi}{\partial t^2} = \frac{\partial^2\chi_1}{\partial t^2} + \frac{\partial^2\chi_2}{\partial t^2}. \tag{47.19}$$

Se sigue que $\partial^2\chi/\partial x^2 = (1/c_s^2)\partial^2\chi/\partial t^2$, por lo que hemos verificado el principio de superposición. La demostración del principio de superposición proviene de que la ecuación de onda es *lineal* en χ.

Podemos esperar ahora que una onda luminosa plana propagándose en la dirección x polarizada de modo que el campo eléctrico esté en la dirección y, satisfaga la ecuación de onda:

$$\frac{\partial^2 E_y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 E_y}{\partial t^2}, \tag{47.20}$$

donde c es la velocidad de la luz. Esta ecuación de onda es una de las consecuencias de las ecuaciones de Maxwell. Las ecuaciones de la electrodinámica llevarán a la ecuación de onda para la luz como las ecuaciones de la mecánica llevan a la ecuación de onda para el sonido.

47-5 La velocidad del sonido

Nuestra deducción de la ecuación de onda para el sonido nos ha dado una *fórmula* que conecta la velocidad de la onda a la rapidez de variación de la presión con la densidad a presión normal,

$$c_s^2 = \left(\frac{dP}{d\rho}\right)_0. \tag{47.21}$$

Ahora bien, para calcular esta rapidez de variación es esencial saber cómo varía la temperatura. En una onda sonora sería de esperar que en la región de compresión la temperatura aumente y que en la región de enrarecimiento la temperatura baje. Newton fue el primero en calcular la rapidez de variación de la presión con la densidad y supuso que la temperatura no variaba. Argumentó que el calor era conducido de una región a otra tan rápidamente que la temperatura no podía elevarse o caer. Este argumento da la velocidad isotérmica del sonido y es erróneo. La deducción correcta fue dada más tarde por Laplace, quien propuso la idea opuesta: que la presión y la temperatura varían adiabáticamente en una onda sonora. Además, el flujo de calor de la región comprimida a la enrarecida es despreciable mientras la longitud de onda sea larga respecto al camino libre medio. Con esta condición la poquísima cantidad de flujo de calor en una onda sonora no afecta la velocidad, aunque da una pequeña absorción de la energía sonora. Es lícito esperar que esta absorción aumenta a medida que la longitud de onda se aproxime al camino libre medio, pero estas longitudes de onda son menores que las del sonido audible en factores de cerca de un millón.

La variación real de la presión con la densidad de una onda sonora es la que no permite flujo de calor. Ésta corresponde a la variación adiabática, la cual ya

vimos que era $PV^\gamma =$ constante, donde V era el volumen. Como la densidad ρ es inversamente proporcional a V, la relación adiabática entre P y ρ es:

$$P = \text{constante}\,\rho^\gamma, \tag{47.22}$$

de la cual obtenemos $dP/d\rho = \gamma P/\rho$. Para la velocidad del sonido tenemos entonces la relación

$$c_s^2 = \frac{\gamma P}{\rho}. \tag{47.23}$$

También podemos escribir $c_s^2 = \gamma PV/\rho V$ y utilizar la relación $PV = NkT$. Además vemos que ρV es la masa del gas, que se puede expresar también como Nm, o como μ por mol, donde m es la masa de una molécula y μ es el peso molecular. En esta forma encontramos:

$$c_s^2 = \frac{\gamma kT}{m} = \frac{\gamma RT}{\mu}, \tag{47.24}$$

según la cual es evidente que la velocidad del sonido depende únicamente de la temperatura del gas y no de la presión o de la densidad. También hemos observado que:

$$kT = \tfrac{1}{3} m\langle v^2\rangle, \tag{47.25}$$

donde $\langle v^2\rangle$ es el promedio de los cuadrados de la velocidad de las moléculas. Por lo tanto, $c_s^2 = (\gamma/3)\langle v^2\rangle$, es decir,

$$c_s = \left(\frac{\gamma}{3}\right)^{1/2} v_{\text{mc}}. \tag{47.26}$$

Esta ecuación establece que la velocidad del sonido vale aproximadamente $1/(3)^{1/2}$ multiplicado por cierta velocidad media, v_{mc} de las moléculas (la raíz cuadrada de la media de los cuadrados de la velocidad). En otras palabras, la velocidad del sonido es del mismo orden de magnitud que la velocidad de las moléculas; en realidad es un poco menor que esta velocidad media.

Naturalmente que ese resultado era de esperar, porque una perturbación como una variación de presión se propaga, después de todo, debido al movimiento de las moléculas. Sin embargo, ese argumento no nos da la velocidad precisa de propagación; podría haber resultado que el sonido fuese transportado principalmente por las moléculas más rápidas, o por las más lentas. Es razonable y satisfactorio que la velocidad del sonido sea aproximadamente $\frac{1}{2}$ de la velocidad molecular media v_{mc}.

48

Pulsaciones

48-1 Sumando dos ondas

Hace algún tiempo discutimos con mucho detalle las propiedades de las ondas de luz y su interferencia —dicho con otras palabras, los efectos de la superposición de dos ondas provenientes de fuentes diferentes—. En todos estos análisis supusimos que las frecuencias de las fuentes eran iguales. En este capítulo discutiremos algunos de los fenómenos que resultan de la interferencia de dos fuentes que tienen frecuencias *diferentes*.

Es fácil imaginarse lo que va a pasar. Procediendo de la misma manera que antes, supongan que tenemos dos fuentes oscilantes iguales de la misma frecuencia cuyas fases están ajustadas de modo tal, digamos, que las señales llegan en fase a algún punto P. En dicho punto, si se trata de luz, la luz es muy intensa; si se trata de sonido, es muy fuerte, o si de electrones, llegan muchos de ellos. Por otro lado, si las señales que llegan estuvieran desfasadas 180° no obtendríamos señal en P, ya que la amplitud total es entonces un mínimo. Supongan ahora que alguien gira el "botón de fase" de una de las fuentes y cambia la fase en P hacia adelante y hacia atrás, digamos, haciéndola primero 0°, luego 180° y así sucesivamente. En este caso encontraríamos, naturalmente, variaciones en la intensidad de la señal total. Pero también vemos que si la fase de una fuente varía lentamente con relación a la de otra de una manera uniforme y gradual, empezando en cero, subiendo a 10, 20, 30, 40°, etc., entonces lo que mediríamos en P sería una serie de pulsaciones fuertes y débiles, puesto que cuando la fase se mueve en 360° la amplitud vuelve a un máximo. Naturalmente, decir que una fuente está variando su fase relativa a otra fuente a una velocidad uniforme, es lo mismo que decir que el número de oscilaciones por segundo es levemente diferente para las dos.

Así, sabemos la respuesta; si tenemos dos fuentes con frecuencias levemente diferentes encontraríamos, como resultado neto, una oscilación con una lenta intensidad pulsante. ¡Esto es, realmente, todo lo que hay respecto a esta materia!

También es muy fácil formular matemáticamente este resultado. Supongan, por ejemplo, que tenemos dos ondas y analizamos simplemente lo que llega a P, sin preocuparnos por el momento de todas las relaciones espaciales. De una fuente, digamos, tendríamos $\cos\omega_1 t$ y de la otra $\cos\omega_2 t$, donde las ω no son exactamente iguales. Naturalmente, las amplitudes podrían no ser iguales

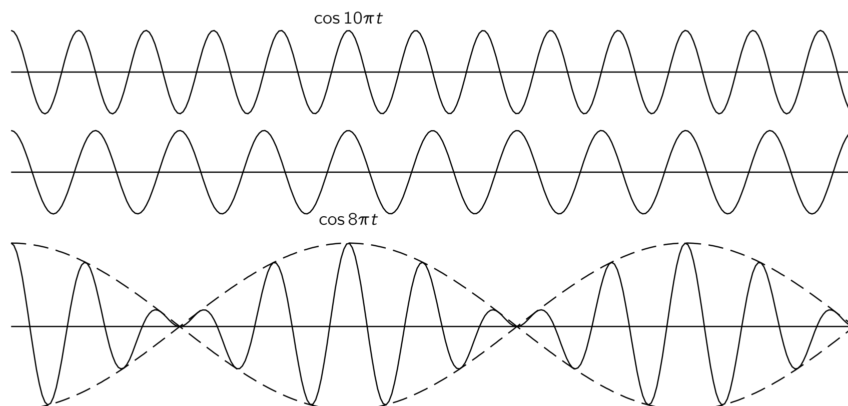


Figura 48-1. Superposición de dos ondas cosenoidales cuyas amplitudes están en la relación 8 : 10. La repetición exacta del diagrama dentro de cada "pulsación", no es típica del caso general.

tampoco, pero podemos resolver el problema general más tarde; tenemos primero el caso en que las amplitudes son iguales. Entonces, la amplitud total en P es la suma de esos dos cosenos. Si representamos las amplitudes de las ondas en función del tiempo, como en la figura 48-1, vemos que donde las crestas coinciden obtenemos una onda fuerte y donde coinciden una cresta y un valle obtenemos prácticamente cero y cuando vuelven a coincidir las crestas obtenemos de nuevo una onda fuerte.

Matemáticamente, necesitamos solamente sumar dos cosenos y arreglar el resultado de alguna forma. Existe una cantidad de relaciones útiles entre los cosenos, que no son difíciles de derivar. Naturalmente, sabemos que:

$$e^{i(a+b)} = e^{ia}e^{ib}, \tag{48.1}$$

y que e^{ia} tiene una parte real, $\cos a$, y una parte imaginaria, $\operatorname{sen} a$. Si tomamos la parte real de $e^{i(a+b)}$, obtenemos $\cos(a+b)$. Si efectuamos la multiplicación

$$e^{ia}e^{ib} = (\cos a + i \operatorname{sen} a)(\cos b + i \operatorname{sen} b),$$

obtenemos $\cos a \cos b - \operatorname{sen} a \operatorname{sen} b$, más algunas partes imaginarias. Pero ahora necesitamos solamente la parte real, o sea,

$$\cos(a+b) = \cos a \cos b - \operatorname{sen} a \operatorname{sen} b. \tag{48.2}$$

Ahora bien, si cambiamos el signo de b, como el coseno no cambia el signo y el seno sí, la misma ecuación para b negativo es:

$$\cos(a-b) = \cos a \cos b + \operatorname{sen} a \operatorname{sen} b. \tag{48.3}$$

Si sumamos estas dos ecuaciones, desaparecen los senos y vemos que el producto de dos cosenos es un medio del coseno de la suma más un medio del coseno de la diferencia:

$$\cos a \cos b = \tfrac{1}{2}\cos(a+b) + \tfrac{1}{2}\cos(a-b). \tag{48.4}$$

Ahora podemos invertir la fórmula y encontrar una para $\cos\alpha + \cos\beta$ si hacemos simplemente $\alpha = a + b$ y $\beta = a - b$. Esto es, $a = \frac{1}{2}(\alpha+\beta)$ y $b = \frac{1}{2}(\alpha-\beta)$, de modo que:

$$\cos\alpha + \cos\beta = 2\cos\tfrac{1}{2}(\alpha+\beta)\cos\tfrac{1}{2}(\alpha-\beta). \tag{48.5}$$

Podemos ahora analizar nuestro problema. La suma de $\cos\omega_1 t$ y $\cos\omega_2 t$ es:

$$\cos\omega_1 t + \cos\omega_2 t = 2\cos\tfrac{1}{2}(\omega_1+\omega_2)t\cos\tfrac{1}{2}(\omega_1-\omega_2)t. \tag{48.6}$$

Pero supongamos que las dos frecuencias son casi iguales, de modo que $\frac{1}{2}(\omega_1+\omega_2)$ es la frecuencia promedio y es, más o menos, igual a ambas. Pero $\omega_1 - \omega_2$ es *mucho menor* que ω_1 o ω_2 porque, como supusimos, ω_1 y ω_2 son casi iguales. Esto significa que podemos representar la solución diciendo que hay una onda cosenoidal de alta frecuencia más o menos como las que teníamos, pero que su "tamaño" varía despacio —su "tamaño" está pulsando con una frecuencia que aparentemente es $\frac{1}{2}(\omega_1-\omega_2)$—. Pero ¿es esta frecuencia a la que se oyen las pulsaciones? Aunque (48.6) dice que la amplitud varía como $\cos\frac{1}{2}(\omega_1-\omega_2)t$, lo que realmente nos está diciendo es que las oscilaciones de alta frecuencia están contenidas entre dos curvas cosenoidales opuestas (las de trazos en la figura 48-1). Basándose en esto se podría decir que la amplitud varía a la frecuencia $\frac{1}{2}(\omega_1-\omega_2)$, pero si hablamos de la *intensidad* de la onda, debemos considerar que tiene el doble de esta frecuencia. Es decir, la modulación de la amplitud, en el sentido de la magnitud de su intensidad, es a la frecuencia $\omega_1 - \omega_2$, aunque la fórmula nos diga que multiplicamos por una onda cosenoidal a mitad de dicha frecuencia. La base técnica de esa diferencia es que la onda de alta frecuencia tiene una relación de fase un poco diferente en el segundo semiciclo.

Ignorando esta pequeña complicación, podemos concluir que si sumamos dos ondas de frecuencias ω_1 y ω_2, obtendremos una onda resultante de frecuencia promedio $\frac{1}{2}(\omega_1 + \omega_2)$ que oscila en intensidad con una frecuencia $\omega_1 - \omega_2$.

Si las dos amplitudes son diferentes, lo podemos volver a hacer todo de nuevo multiplicando los cosenos por amplitudes diferentes A_1 y A_2 y hacer muchas operaciones, reordenando, etc., usando ecuaciones como (48.2)-(48.5). Sin embargo, hay otras maneras más fáciles de efectuar el mismo análisis. Por ejemplo, sabemos que es mucho más fácil trabajar con exponenciales que con senos y cosenos y que podemos representar $A_1 \cos \omega_1 t$ como la parte real de $A_1 e^{i\omega_1 t}$. La otra onda sería análogamente la parte real de $A_2 e^{i\omega_2 t}$. Si las sumamos, obtenemos $A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t}$. Si sacamos el factor común a la frecuencia promedio, resulta:

$$A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t} = e^{i(\omega_1+\omega_2)t/2}[A_1 e^{i(\omega_1-\omega_2)t/2} + A_2 e^{-i(\omega_1-\omega_2)t/2}]. \qquad (48.7)$$

Tenemos de nuevo la onda de alta frecuencia con una modulación a la frecuencia más baja.

48-2 Notas pulsadas y modulación

Si nos preguntan ahora por la intensidad de la onda de la ecuación (48.7), podemos tomar o el módulo al cuadrado del primer miembro o el del segundo. Tomemos el del primer miembro. La intensidad es entonces:

$$I = A_1^2 + A_2^2 + 2A_1 A_2 \cos(\omega_1 - \omega_2)t. \qquad (48.8)$$

Vemos que la intensidad se infla y se desinfla a una frecuencia $\omega_1 - \omega_2$, variando entre los límites $(A_1 + A_2)^2$ y $(A_1 - A_2)^2$. Si $A_1 \neq A_2$, la intensidad mínima no es cero.

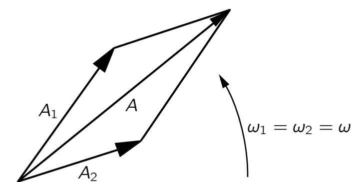


Figura 48-2. Resultante de dos vectores complejos de igual frecuencia.

Otra manera de representar esta idea es mediante un dibujo, como el de la figura 48-2. Dibujando un vector de longitud A_1 que rota a una frecuencia ω_1 para representar una de las ondas en el plano complejo. Dibujamos otro vector de longitud A_2 que rota a una frecuencia ω_2 para representar la segunda onda. Si las dos frecuencias son exactamente iguales, su resultado tiene una longitud fija y sigue rotando, y obtenemos de las dos una intensidad fija y definida. Pero si las frecuencias son levemente diferentes, los dos vectores complejos rotan con velocidades diferentes. La figura 48-3 muestra a qué se parece esta situación con relación al vector $A_1 e^{i\omega_1 t}$. Vemos que A_2 gira apartándose lentamente de A_1 y así la amplitud que obtenemos sumándolos es primero fuerte y luego, a medida que se aleja, cuando la posición relativa es de 180°, la resultante se hace particularmente débil, y así sucesivamente. A medida que los vectores rotan, la amplitud del vector suma se hace mayor y menor, y de este modo la intensidad pulsa. Es una idea relativamente simple y hay muchos modos diferentes de representarla.

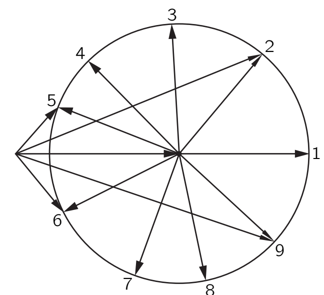


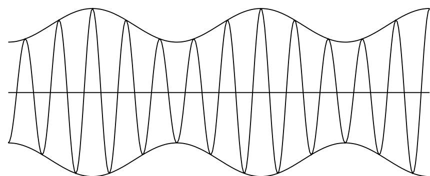
Figura 48-3. Resultante de dos vectores complejos de frecuencia desigual, vista desde el sistema de referencia de un vector que rota. Se muestran nueve posiciones sucesivas del vector que rota lentamente.

Es muy fácil observar experimentalmente el efecto. En el caso acústico, podemos disponer dos altoparlantes accionados por dos osciladores individuales, uno por cada altoparlante, de modo que cada uno de ellos da un tono. Así, recibimos una nota de una fuente y otra nota diferente de la otra fuente. Si hacemos que las frecuencias sean exactamente iguales, el efecto resultante tendrá una intensidad definida en una posición dada del espacio. Si entonces los desintonizamos un poquito, apreciamos ciertas variaciones en la intensidad. Cuanto más los desintonizamos, más rápidas son las variaciones de sonido. El oído tiene alguna dificultad en seguir variaciones más rápidas que unas 10 por segundo.

También podemos observar el efecto en un osciloscopio que desdobla simplemente la suma de las corrientes que entran a los dos altoparlantes. Si la frecuencia de pulsación es relativamente baja, vemos sencillamente un tren de ondas sinusoidal cuya amplitud pulsa, pero cuando hacemos las pulsaciones más rápidas vemos la clase de onda que muestra la figura 48-1. Cuando hacemos las diferencias de frecuencia mayores, las "jorobas" se mueven más juntas. También, si las amplitudes no son iguales y hacemos una señal más fuerte que la otra, entonces obtenemos una onda cuya amplitud no se hace nunca cero, que es

exactamente lo que esperamos. Todo marcha como es debido, tanto acústica como eléctricamente.

¡El fenómeno opuesto también se da! En radiotransmisión, usando la llamada *amplitud modulada* (AM), la estación de radio emite el sonido como sigue: el radiotransmisor tiene una oscilación eléctrica CA a una frecuencia muy alta, por ejemplo 800 kilociclos por segundo, en la banda de emisión. Si se pone a funcionar esta *señal portadora*, la estación de radio emite una onda de amplitud uniforme a 800 000 oscilaciones por segundo. La manera de transmitir la "información", el tipo inútil de información acerca de qué clase de automóvil comprar, es que cuando alguien habla por el micrófono, la amplitud de la señal portadora cambia al compás de las vibraciones acústicas que entran en el micrófono.


Figura 48-4. Una onda portadora modulada. En este dibujo esquemático, $\omega_c/\omega_m = 5$. En una onda real de radio, $\omega_c/\omega_m \sim 100$.

Si tomamos como el caso matemático más simple la situación en que una soprano está dando una nota perfecta, con oscilaciones sinusoidales perfectas de sus cuerdas vocales, obtenemos entonces una señal cuya intensidad varía como muestra la figura 48-4. Además, la alternación de la audiofrecuencia se recobra entonces en el receptor, nos libramos de la onda portadora y solamente consideramos la envolvente, que representa las oscilaciones de las cuerdas vocales o sonido de la cantante. El altoparlante hace vibraciones correspondientes a la misma frecuencia en el aire y el oyente es esencialmente incapaz de distinguir la diferencia, eso es lo que dicen. Debido a un número de distorsiones y otros efectos sutiles es posible, en realidad, decir si estamos escuchando la radio o una soprano en persona; de todos modos la idea es la que hemos indicado.

48-3 Bandas laterales

Matemáticamente, la onda modulada descrita anteriormente se expresaría en la forma:

$$S = (1 + b\cos\omega_m t)\cos\omega_c t, \tag{48.9}$$

donde ω_c representa la frecuencia de la portadora y ω_m es la frecuencia de audio. De nuevo usamos todos los teoremas de los cosenos, o podemos usar $e^{i\theta}$; da igual —es más fácil con $e^{i\theta}$, pero es lo mismo—. Tenemos entonces:

$$S = \cos\omega_c t + \tfrac{1}{2}b\cos(\omega_c + \omega_m)t + \tfrac{1}{2}b\cos(\omega_c - \omega_m)t. \tag{48.10}$$

Así, desde otro punto de vista podemos decir que la onda de salida del sistema consiste en tres ondas superpuestas: primero, la onda regular de frecuencia ω_c, es decir, frecuencia de la portadora y luego dos ondas nuevas a dos frecuencias nuevas. Una es la frecuencia portadora más la frecuencia de modulación y la otra es la frecuencia portadora menos la frecuencia de modulación. Por lo tanto, si hacemos una especie de representación de la intensidad generada por el generador en función de la frecuencia, encontraríamos una gran cantidad de intensidad a la frecuencia de la portadora, naturalmente, pero cuando un cantante empieza a actuar también encontraríamos de repente intensidad proporcional a la intensidad del cantante, b^2, a frecuencias $\omega_c + \omega_m$ y $\omega_c - \omega_m$, como muestra la figura 48-5. Esto es lo que se llama *bandas laterales*; cuando hay una señal modulada proveniente del transmisor hay bandas laterales. Si hay más de una nota al mismo tiempo, digamos ω_m y $\omega_{m'}$, si hay dos instrumentos tocando, o si hay cualquier otra onda complicada con cosenos, entonces por supuesto podemos describir matemáticamente que obtenemos más ondas que corresponden a las frecuencias $\omega_c \pm \omega_{m'}$.

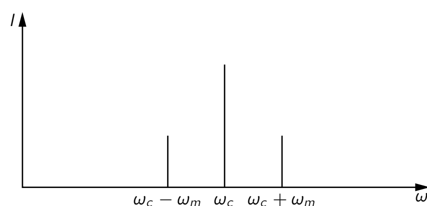


Figura 48-5. Espectro de frecuencia de una onda portadora ω_c modulada por una única onda cosenoidal ω_m.

Por lo tanto, cuando hay una modulación complicada que se puede representar como la suma de varios cosenos,* encontramos que el transmisor real está

* Una pequeña nota al margen. ¿En qué circunstancias se puede representar una curva como suma de un gran número de cosenos? *Respuesta*: en todas las circunstancias ordinarias excepto para ciertos casos que los matemáticos pueden inventar. Naturalmente, la curva debe tener solamente un valor en un punto dado, y no debe ser una curva loca que salta un número infinito de veces en una distancia infinitesimal, o algo parecido. Pero aparte de estas restricciones, cualquier curva razonable (una que un cantante es capaz de hacer agitando sus cuerdas vocales) siempre se puede componer sumando ondas cosenoidales.

transmitiendo en un intervalo de frecuencias, es decir, la frecuencia portadora más o menos la frecuencia máxima que contiene la señal de modulación.

Aunque al principio podamos pensar que un radiotransmisor transmite solamente a la frecuencia nominal de la portadora, puesto que hay grandes y superestables osciladores de cristal y todo está ajustado para que sean precisamente 800 kilociclos, desde el momento en que alguien *anuncie* que están a 800 kilociclos, modula los 800 kilociclos y por consiguiente ¡ya no están precisamente a 800 kilociclos! Ahora bien, supongan que amplificadores están construidos en tal forma que transmitan en un buen intervalo de la sensibilidad del oído (el oído puede percibir hasta 20 000 ciclos por segundo, pero corrientemente los transmisores y receptores de radio no trabajan más allá de 10 000, por lo que no oímos las partes más altas), entonces, cuando un hombre habla, su voz puede contener frecuencias más altas, digamos, que 10 000 ciclos, por lo que el transmisor está transmitiendo frecuencias que pueden ir desde 790 hasta 810 kilociclos por segundo. Ahora bien, si hubiese otra estación a 795 kc/s, habría mucha confusión. También, si hiciéramos nuestro receptor tan sensible que solamente tomase 800 y no los 10 kilociclos de ambos lados, no oiríamos lo que el hombre dijese, ¡ya que la información estaría en esas otras frecuencias! Por lo tanto, es absolutamente esencial colocar las estaciones algo separadas para que sus bandas laterales no se superpongan y además el receptor no debe ser tan selectivo que no permita la recepción de las bandas laterales así como de la frecuencia nominal principal. En el caso del sonido, este problema no causa realmente mucha dificultad. Podemos oír en un rango de ± 20 kc/s y corrientemente tenemos desde 500 hasta 1 500 kc/s en la banda de emisión, por lo que hay lugar suficiente para muchas estaciones.

El problema de la televisión es más difícil. Cuando el haz electrónico atraviesa la carga del tubo de imagen, hay varias manchas pequeñas de luz y sombra. Esta "luz" y "sombra" es la "señal". Ahora bien, el haz barre de ordinario toda la imagen, 500 líneas, en una trigésima parte de segundo aproximadamente. Consideremos que la resolución de la imagen vertical y horizontalmente es más o menos igual, por lo que hay el mismo número de manchas por centímetro según una línea de barrido. Queremos ser capaces de distinguir sombra de luz, en 500 líneas, digamos. Para que podamos hacer esto con ondas cosenoidales, la más pequeña longitud de onda que se necesita corresponde entonces a una longitud de onda, de máximo a máximo, de la doscientoscincuentava parte del tamaño de la pantalla. De este modo tenemos $250 \times 500 \times 30$ informaciones por segundo. La frecuencia más alta que vamos a transportar, por lo tanto, es próxima a 4 megaciclos por segundo. En la realidad, para conservar las estaciones de televisión independientes tenemos que usar un poquito más que esto. Alrededor de 6 mc/s; parte se usa para llevar la señal de sonido y otra información. Por lo que los canales de televisión tienen seis megaciclos por segundo de anchura. Ciertamente no sería posible transmitir TV en una portadora de 800 kc/s, ya que no podemos modular a una frecuencia más alta que la portadora.

De cualquier modo, la banda de televisión comienza en 54 mc. El primer canal de transmisión, que es el canal 2 (!), tiene un intervalo de frecuencias desde 54 hasta 60 mc/s, que da 6 mc/s de anchura. "Pero —podría decir alguien— acabamos de probar que había bandas laterales a ambos lados, y por lo tanto debería tener anchura doble." Resulta que los ingenieros de radio son inteligentes. Si analizamos la señal de modulación usando términos en senos y cosenos en lugar de sólo cosenos, para permitir diferencias de fase, vemos que hay una relación invariante definida entre la banda lateral en el lado de la alta frecuencia y la banda lateral en el lado de la baja frecuencia. Queremos decir que no hay información nueva en la otra banda lateral. Así, lo que se hace es suprimir una de éstas y el receptor está conectado dentro en una forma tal que la información omitida se reconstituye, examinando la única banda lateral y la portadora. La transmisión de banda lateral única es un esquema inteligente para disminuir los anchos de banda necesarios para transmitir información.

48-4 Trenes de ondas localizados

El próximo tema que discutiremos es la interferencia de ondas en el espacio y en el tiempo. Supongan que tenemos dos ondas viajando en el espacio. Sabemos, naturalmente, que podemos representar una onda que viaja en el espacio mediante $e^{i(\omega t-kx)}$. Esto podría ser, por ejemplo, el desplazamiento de una onda de sonido. Además es una solución de la ecuación de onda siempre que $\omega^2 = k^2c^2$, donde c es la velocidad de propagación de la onda. En este caso lo podemos escribir en la forma $e^{-ik(x-ct)}$, que es de la forma general $f(x-ct)$. Por lo tanto, ésta debe ser una onda que está viajando a esta velocidad, ω/k, que es c y todo está bien.

Ahora queremos sumar dos de estas ondas. Supongan que tenemos una onda que está viajando con una frecuencia y otra onda viajando con otra frecuencia. Dejamos al lector la consideración del caso en que las amplitudes son diferentes, no hay diferencia sustancial. Nosotros queremos efectuar $e^{i(\omega_1 t-k_1x)} + e^{i(\omega_2 t-k_2x)}$. Podemos hacerlo usando la misma matemática que cuando sumábamos ondas de señal. Naturalmente que si c es la misma para ambas esto es fácil, puesto que es idéntico a lo que hicimos antes:

$$e^{i\omega_1(t-x/c)} + e^{i\omega_2(t-x/c)} = e^{i\omega_1 t'} + e^{i\omega_2 t'}, \tag{48.11}$$

excepto que ahora la variable es $t' = t - x/c$ en lugar de t. Así, obtenemos la misma clase de modulaciones, pero vemos, naturalmente, que estas modulaciones se están trasladando con la onda. En otras palabras, si sumásemos dos ondas y estas ondas no sólo estuviesen oscilando sino también moviéndose en el espacio, entonces la onda resultante también se trasladaría en el espacio a la misma velocidad.

Ahora querríamos generalizar esto al caso de ondas en las que la relación entre la frecuencia y el número de ondas k no sea tan simple. Ejemplo: un material que tenga un índice de refracción. Ya estudiamos la teoría del índice de refracción en el capítulo 31, donde encontramos que podíamos escribir $k = n\omega/c$, siendo n el índice de refracción. Como ejemplo interesante, encontramos que para los rayos el índice n es:

$$n = 1 - \frac{Nq_e^2}{2\epsilon_0 m\omega^2}. \tag{48.12}$$

Dedujimos realmente una fórmula más complicada en el capítulo 31, pero como ejemplo ésta es tan buena como cualquiera.

Entre paréntesis, sabemos que aun cuando ω y k no son proporcionales linealmente, el cociente ω/k es ciertamente la velocidad de propagación para una frecuencia y número de ondas particulares. Llamamos *velocidad de fase* a este cociente; es la velocidad a la que la fase, o los nodos de una única onda, se trasladarían:

$$v_p = \frac{\omega}{k}. \tag{48.13}$$

Esta velocidad de fase, para el caso de los rayos X en vidrio, es mayor que la velocidad de la luz en el vacío [ya que n en (48.12) es menor que 1], y esto es un poco incómodo, porque ¡no creemos que se pueda enviar señales más rápido que la velocidad de la luz!

Lo que vamos a discutir ahora es la interferencia de dos ondas en las que ω y k tienen una fórmula definida que las relaciona. La fórmula anterior para n dice que k está dado como una función definida de ω. Para ser precisos, en este problema particular la fórmula para k en función de ω es:

$$k = \frac{\omega}{c} - \frac{a}{\omega c}, \tag{48.14}$$

donde $a = Nq_e^2/2\epsilon_0 m$ es una constante. De cualquier modo, para cada frecuencia hay un número de onda definido y queremos sumar dos de estas ondas.

Hagámoslo exactamente como hicimos en la ecuación (48.7):

$$e^{i(\omega_1 t-k_1 x)}+e^{i(\omega_2 t-k_2 x)}=e^{i[(\omega_1+\omega_2)t-(k_1+k_2)x]/2}$$
$$\times\left\{e^{i[(\omega_1-\omega_2)t-(k_1-k_2)x]/2}+e^{-i[(\omega_1-\omega_2)t-(k_1-k_2)x]/2}\right\}. \qquad (48.15)$$

Tomemos de nuevo una onda modulada, una onda que viaja con la frecuencia promedio y el número de onda promedio, pero cuya intensidad está variando con una forma que depende de la diferencia de frecuencia y de la diferencia del número de onda.

Tomemos ahora el caso en que la diferencia entre las dos ondas es relativamente pequeña. Supongamos que estamos sumando dos ondas cuyas frecuencias son casi iguales, entonces $(\omega_1+\omega_2)/2$ es prácticamente igual a cualquiera de las dos ω y lo mismo ocurre con $(k_1+k_2)/2$. Entonces, la velocidad de la onda, de las oscilaciones rápidas, de los nodos, es esencialmente ω/k. Pero, ojo, ¡la velocidad de propagación de la modulación no es la misma! ¿Cuánto tenemos que cambiar x para dar cuenta de cierta cantidad de t? La velocidad de esta onda de modulación es el cociente:

$$v_M=\frac{\omega_1-\omega_2}{k_1-k_2}. \qquad (48.16)$$

La velocidad de modulación se llama algunas veces *velocidad de grupo*. Si tomamos el caso en que la diferencia de la frecuencia es relativamente pequeña y la diferencia de número de onda es, entonces, también relativamente pequeña, esta expresión en el límite tiende a:

$$v_g=\frac{d\omega}{dk}. \qquad (48.17)$$

En otras palabras, para la modulación más lenta, para las pulsaciones más lentas hay una velocidad definida a la que viajan que no es igual a la velocidad de fase de las ondas —¡que cosa tan misteriosa!—.

La velocidad de grupo es la derivada de ω respecto a k y la velocidad de fase es ω/k.

Veamos si podemos entender el por qué. Consideren dos ondas, de nuevo con longitudes de ondas levemente diferentes, como en la figura 48-1. Están desfasadas, en fase, desfasadas, y así sucesivamente. Pero estas ondas representan, realmente, ondas en el espacio que viajan con frecuencias levemente diferentes también. Ahora bien, ya que la velocidad de fase, velocidad de los nodos de estas dos ondas, no es precisamente la misma, sucede algo nuevo. Supongan que cabalgamos sobre una de las ondas y miramos la otra, si ambas fuesen a la misma velocidad, entonces la otra onda permanecería con la misma posición respecto a nosotros, según cabalgamos en la cresta. Cabalgamos en esta cresta y justo frente a nosotros vemos una cresta, si las dos velocidades son iguales las dos crestas permanecen una encima de otra. Pero *no es así*, las dos velocidades no son realmente iguales. Hay solamente una pequeña diferencia de frecuencia y por lo tanto solamente una pequeña diferencia de velocidad, pero a causa de esta diferencia de velocidad, a medida que cabalgamos, la otra onda se mueve despacio hacia adelante, digamos, o hacia atrás, con relación a nuestra onda. Así, según pasa el tiempo, ¿qué sucede al modo? Si movemos un tren de ondas sólo un ápice hacia adelante, el nodo se mueve hacia adelante (o hacia atrás) una distancia considerable. Es decir, la suma de estas dos ondas tiene una envolvente, y a medida que las ondas avanzan, la envolvente cabalga sobre ellas a una velocidad diferente. La *velocidad de grupo* es la velocidad a la que se transmitirían las señales moduladas.

Si hiciéramos una señal, es decir, una especie de cambio de onda que se pudiera reconocer cuando se le escucha, una especie de modulación, entonces ésta viajaría con la velocidad de grupo, siempre que las modulaciones fueran relativamente lentas. (Cuando son rápidas, es mucho más difícil hacer el análisis.)

Ahora podemos mostrar (por fin) que la velocidad de propagación de los rayos X en un bloque de carbono *no* es mayor que la velocidad de la luz, aunque la velocidad de fase *es* mayor que la velocidad de la luz. Para hacer esto, debemos encontrar $d\omega/dk$ que obtenemos derivando (48.14): $dk/d\omega = 1/c + a/\omega^2 c$. La velocidad de grupo, por lo tanto, es la inversa, o sea,

$$v_g = \frac{c}{1 + a/\omega^2}, \tag{48.18}$$

¡que es menor que c! Así, aunque las fases pueden viajar más rápido que la velocidad de la luz, las señales de modulación viajan más despacio, y ¡ésta es la solución de la paradoja aparente! Naturalmente que si tenemos el caso simple en que $\omega = kc$, entonces $d\omega/dk$ es también c. Por lo tanto, cuando todas las fases tienen la misma velocidad, el grupo tiene por supuesto la misma velocidad.

48-5 Amplitudes de probabilidad para partículas

Consideremos ahora un ejemplo más de la velocidad de fase que es en extremo interesante. Tiene que ver con la mecánica cuántica. Sabemos que la amplitud de encontrar una partícula en el lugar puede variar, en algunas circunstancias, en el espacio y en el tiempo, digamos que en una dimensión, en la siguiente forma:

$$\psi = Ae^{i(\omega t - kx)}, \tag{48.19}$$

donde ω es la frecuencia, que se relaciona con la idea clásica de energía mediante $E = \hbar\omega$, y k es el número de onda, que se relaciona con el momentum mediante $p = \hbar k$. Diríamos que la partícula tiene un momentum definido p si el número de onda fuese exactamente k, esto es, una onda perfecta que marcha con la misma amplitud en todos los lugares. La ecuación (48.19) da la amplitud y si tomamos el cuadrado del módulo, obtenemos la probabilidad relativa de encontrar la partícula en función de la posición y del tiempo. Es una *constante*, que indica que la probabilidad de encontrar una partícula en cualquier lugar es la misma. Supongan ahora, por el contrario, que tenemos una situación donde sabemos que es más probable que la partícula esté en un lugar que en otro. Representaríamos tal situación por una onda que tiene un máximo y se anula por ambos lados (figura 48-6). [No es lo mismo que a una onda como (48-1) que tiene una serie de máximos, pero es posible, agregando varias ondulaciones de casi lo mismo de ω y k juntos, para deshacerse de todos menos un máximo.]

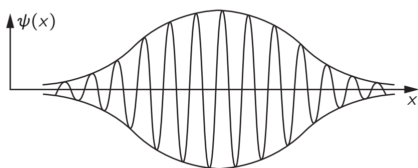


Figura 48-6. Un tren de ondas localizado.

Ahora bien, en estas circunstancias, puesto que el cuadrado de (48.19) representa la probabilidad de encontrar una partícula en algún lugar, sabemos que en un instante dado es más probable que la partícula se encuentre cerca del centro del "montón" donde la amplitud de la onda es máxima. Si esperamos ahora unos instantes, las ondas se moverán y después de algún tiempo el "montón" estará en algún otro lugar. Si supiésemos que la partícula originalmente estaba situada en algún lugar, clásicamente *esperaríamos* que más tarde estuviese en otro lugar como cosa sabida, porque después de todo tiene una *velocidad* y un momentum. La teoría cuántica se reducirá entonces a la teoría correcta para la relación de momentum, energía y velocidad, solamente si la velocidad de grupo, velocidad de la modulación, es igual a la velocidad que obtendríamos clásicamente para una partícula de idéntico momentum.

Ahora es necesario demostrar si éste es o no el caso. Según la teoría clásica, la energía se relaciona con la velocidad mediante:

$$E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}. \tag{48.20}$$

Del mismo modo, el momentum es:

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}}. \tag{48.21}$$

Ésta es la teoría clásica y como consecuencia de ella, eliminando v, podemos ver:

$$E^2 - p^2c^2 = m^2c^4.$$

Éste es el gran resultado de dimensión cuatro del que hemos estado hablando y mencionando que $p_\mu p_\mu = m^2$; ésta es la relación entre energía y momentum en la teoría clásica. Pero esto significa, puesto que estas E y p van a ser ω y k, sustituyendo $E = \hbar\omega$ y $p = \hbar k$, que para la mecánica cuántica es necesario que:

$$\frac{\hbar^2\omega^2}{c^2} - \hbar^2 k^2 = m^2c^2. \tag{48.22}$$

Ésta es, pues, la relación entre la frecuencia y el número de onda de una onda cuántica de amplitud que representa una partícula de masa m. De esa ecuación podemos deducir que ω es:

$$\omega = c\sqrt{k^2 + m^2c^2/\hbar^2}.$$

¡Aquí de nuevo la velocidad de fase, ω/k, es mayor que la velocidad de la luz!

Consideremos ahora la velocidad de grupo. La velocidad de grupo debería ser $d\omega/dk$, velocidad a la que se mueven las modulaciones. Tenemos que derivar una raíz cuadrada, lo cual no es muy difícil. La derivada es:

$$\frac{d\omega}{dk} = \frac{kc}{\sqrt{k^2 + m^2c^2/\hbar^2}}.$$

Pero la raíz cuadrada es después de todo ω/c, por lo que podemos escribir esto en la forma $d\omega/dk = c^2k/\omega$. Más aún k/ω es p/E, por lo que

$$v_g = \frac{c^2p}{E}.$$

Pero de acuerdo con (48.20) y (48.21), $c^2p/E = v$, velocidad de la partícula según la mecánica clásica. Así, vemos que la relación cuántica fundamental $E = \hbar\omega$ y $p = \hbar/k$ para la identificación de ω y k con los E y p clásicos, solamente produce la ecuación $w^2 - k^2c^2 = m^2c^4/\hbar^2$, mientras que ahora también entendemos las relaciones (48.20) y (48.21) que relacionan E y p con la velocidad. La velocidad de grupo, naturalmente, debe ser la velocidad de la partícula si es que la interpretación va a tener algún sentido. Si pensamos que la partícula está aquí en un instante y luego 10 minutos más tarde pensamos que está allá, como dijo la mecánica cuántica, la distancia recorrida por el “montón” dividida entre el intervalo de tiempo debe ser, clásicamente, la velocidad de la partícula.

48-6 Ondas en tres dimensiones

Concluiremos ahora nuestra discusión de las ondas con unas pocas observaciones generales acerca de la ecuación de onda. Se pretende con estas observaciones dar una visión del futuro —no es para que podamos entender todo exactamente ahora mismo, sino más bien ver qué aspecto van a tener las cosas cuando estudiemos ondas un poco más—. Ante todo, la ecuación de onda para el sonido en una dimensión era:

$$\frac{\partial^2\chi}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2},$$

donde c es la velocidad de cualquier cosa que sea la onda —en el caso del sonido, es la velocidad del sonido; en el caso de la luz, es la velocidad de la luz—. Demostramos que para una onda de sonido los desplazamientos se propagarían a una cierta velocidad. Pero el exceso de presión también se propaga a una cierta velocidad y lo mismo ocurre con el exceso de densidad. Por ello, deberíamos esperar que la presión satisficiera la misma ecuación, como ocurre en realidad. Se lo dejaremos al lector para que lo demuestre. *Sugerencia*: ρ_e es proporcional a la rapidez de variación de χ respecto a x. Por lo que si derivamos la ecuación de onda respecto a x, descubriremos inmediatamente que $\partial\chi/\partial x$ satisface la misma ecuación. Es decir, que ρ_e satisface la misma ecuación. Pero P_e es proporcional a ρ_e y en consecuencia P_e también la satisface. En una palabra, la presión, los desplazamientos, todo, satisface la misma ecuación de onda.

Generalmente se ve la ecuación de onda para sonido escrita en términos de presión en lugar de en función de desplazamiento, porque la presión es un escalar y no tiene dirección. Pero, el desplazamiento es un vector y tiene dirección y entonces resulta más fácil analizar la presión.

El próximo punto a discutir tiene que ver con la ecuación de onda en tres dimensiones. Sabemos que la solución para una onda sonora en una dimensión es $e^{i(\omega t-kx)}$, con $\omega = kc_s$, pero también sabemos que en tres dimensiones una onda estaría representada por $e^{i(\omega t-k_x x-k_y y-k_z z)}$, donde en este caso $\omega^2 = k^2 c_s$, que es, naturalmente, $(k_x^2+k_y^2+k_z^2)c_s^2$. Ahora lo que queremos hacer es conjeturar cuál es la ecuación de onda correcta en tres dimensiones. Para el caso del sonido esto se puede deducir, naturalmente, a través del mismo razonamiento dinámico en tres dimensiones que hicimos en una dimensión. Pero no haremos esto; en su lugar, escribimos solamente el resultado: la ecuación para la presión (o desplazamiento, o lo que sea) es

$$\frac{\partial^2 P_e}{\partial x^2} + \frac{\partial^2 P_e}{\partial y^2} + \frac{\partial^2 P_e}{\partial z^2} = \frac{1}{c_s^2}\,\frac{\partial^2 P_e}{\partial t^2}. \tag{48.23}$$

Podemos ver que esto es verdad sustituyendo $e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}$. Es claro que cada vez que derivamos respecto a x, multiplicamos por $-ik_x$. Si derivamos dos veces, es equivalente a multiplicar por $-k_x^2$, por lo que el primer término se convertiría en $-k_x^2 P_e$, para esa onda. Del mismo modo, el segundo término se hace $-k_y^2 P_e$, y el tercero $-k_z^2 P_e$. A la derecha obtenemos $-(\omega^2/c_s^2)P_e$. Entonces, si simplificamos las P_e y cambiamos el signo, la relación entre k y ω es la que nosotros queremos.

Haciendo también las cosas hacia atrás, no podemos resistir la tentación de escribir la gran ecuación que corresponde a la ecuación de dispersión* (48.22) para ondas cuánticas. Si ϕ representa la amplitud de encontrar una partícula en la posición x, y, z, en el tiempo t, entonces la gran ecuación de la mecánica cuántica para partículas libres es ésta:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} - \frac{1}{c^2}\,\frac{\partial^2 \phi}{\partial t^2} = \frac{m^2c^2}{\hbar^2}\,\phi. \tag{48.24}$$

En primer lugar, el carácter relativista de esta expresión está sugerido por la aparición de x, y, z y t en la combinación habitual que la relatividad usa generalmente. En segundo lugar, es una ecuación de onda que, si probamos una onda plana, produciría como consecuencia que $-k^2 + \omega^2/c^2 = m^2c^2/\hbar^2$, que es la relación correcta para la mecánica cuántica. Hay aún otra gran cosa contenida en la ecuación de onda, el hecho de que cualquier superposición de ondas es también una solución. Por lo tanto, esta ecuación contiene todo lo que hemos estado discutiendo hasta ahora de mecánica cuántica y relatividad, ¡por lo menos siempre que se trate de una única partícula en el espacio vacío sin potenciales o fuerzas externas en él!

48-7 Modos normales de vibración

Volvemos ahora a otro ejemplo del fenómeno de pulsaciones que es bastante curioso y un poco diferente. Imaginen dos péndulos iguales que tienen, entre ellos, una conexión elástica bastante débil. Han sido construidos en lo posible de la misma longitud. Si apartamos uno y lo soltamos, se mueve hacia adelante y hacia atrás y cuando realiza ese movimiento tira del resorte de conexión, por lo que realmente es una máquina para generar una fuerza que tiene la frecuencia natural del otro péndulo. Además, como consecuencia de la teoría de la resonancia, que ya estudiamos, cuando aplicamos una fuerza sobre algo a la frecuencia apropiada, lo arrastrará. Por lo que, con toda seguridad, un péndulo moviéndose hacia adelante y hacia atrás arrastrará el otro. Sin embargo, en este caso sucede algo

* N. del T. Dispersión está usada aquí en un sentido análogo a la dispersión de un prisma, no debe confundirse con la dispersión de ondas o partículas al chocar contra un objeto. O sea, que estrictamente dispersión en este caso quiere decir el fenómeno por el cual la velocidad de fase de una onda depende de su frecuencia.

nuevo, ya que la energía total del sistema es finita y así, cuando un péndulo da parte de su energía al otro para arrastrarlo, se encuentra a sí mismo perdiendo energía hasta que, si hay una sincronización apropiada con la velocidad, pierde toda su energía y ¡pasa a un estado estacionario! Entonces, es el otro péndulo el que tiene toda la energía y el primero no tiene nada, y según pasa el tiempo vemos que todo funciona en sentido opuesto y que la energía se devuelve al primero; éste es un fenómeno muy interesante y divertido. Dijimos, sin embargo, que está relacionado con la teoría de las pulsaciones y ahora debemos explicar cómo podemos analizar este movimiento desde el punto de vista de dicha teoría.

Observamos que el movimiento de cualquiera de las dos esferitas es una oscilación con una amplitud que varía cíclicamente. Por lo tanto, el movimiento de una de las esferitas se puede analizar presumiblemente en una forma diferente, como suma de dos oscilaciones presentes al mismo tiempo pero que tienen dos frecuencias levemente diferentes. Por lo tanto, debería ser posible encontrar otros dos movimientos en este sistema y sostener que lo que vimos era una superposición de dos soluciones ya que éste es, por supuesto, un sistema lineal. Verdaderamente, es fácil encontrar dos maneras en las que podría empezar el movimiento, siendo cada una un movimiento perfecto de una sola frecuencia —absolutamente periódico—. El movimiento con el que empezamos antes no era estrictamente periódico, puesto que no duró, pronto una esferita pasaba energía a la otra y así cambiaba su amplitud. Pero hay maneras de comenzar el movimiento para que nada cambie y, naturalmente, tan pronto como lo vemos entendemos por qué. Por ejemplo, si hiciésemos que ambos péndulos marchasen juntos, entonces, puesto que ambos son de la misma longitud y el resorte en este caso no hace nada, continuarían por supuesto oscilando para siempre, suponiendo que no hay roce y que todo es perfecto. Por otro lado, hay otro movimiento posible que también tiene una frecuencia definida; esto es, si movemos opuestamente los péndulos separándolos la misma distancia exactamente estarían de nuevo en movimiento absolutamente periódico. Podemos apreciar que el resorte añade solamente un poco a la fuerza restauradora que proporciona la gravedad, eso es todo, y el sistema sigue justamente oscilando a una frecuencia levemente más alta que en el primer caso. ¿Por qué más alta? Porque el resorte está tirando, además de la gravedad, y hace que el sistema sea un poco "más rígido", por lo que la frecuencia de su movimiento es sólo una pizca más alta que la del otro.

Este sistema, pues, tiene dos maneras de oscilar sin variar de amplitud, o bien, puede oscilar de una manera tal que ambos péndulos sigan el mismo camino u oscilen todo el tiempo a una frecuencia o podrían ir en direcciones opuestas a una frecuencia levemente más alta.

Ahora bien, el movimiento real del sistema, ya que es lineal, se puede representar como una superposición de dos. (Recuerden que el tema de este capítulo es los efectos de sumar dos movimientos con frecuencias diferentes.) Por ello, piensen lo que sucedería si combinásemos estas dos soluciones. Si a $t = 0$ los dos movimientos han empezado con amplitudes iguales y en la misma fase, la suma de los dos movimientos significa que una esferita, habiéndosele dado un sentido por el primer movimiento y el otro sentido por el segundo movimiento, está a cero, mientras que la otra esferita, habiendo sido desplazada en el mismo sentido por los dos movimientos, tiene una gran amplitud. A medida que pasa el tiempo, sin embargo, los dos *movimientos* básicos actúan independientemente, por tanto la fase de uno relativa al otro se está corriendo lentamente. Esto significa, que después de un tiempo muy largo, cuando el tiempo es suficiente para que un movimiento pueda haber realizado "$900\frac{1}{2}$" oscilaciones mientras el otro realiza solamente "900", la fase relativa estaría justamente invertida con relación a como estaba antes. Esto es, el movimiento de gran amplitud se habrá hecho cero, y mientras tanto, naturalmente, la esferita inicialmente sin movimiento ¡habrá alcanzado intensidad completa!

Así, vemos que podríamos analizar este movimiento complicado, o bien mediante la idea de que hay resonancia y que uno entrega energía al otro o mediante la superposición de dos movimientos de amplitudes constantes a dos frecuencias diferentes.

49

Modos de vibración

49-1 Reflexión de ondas

Este capítulo considerará algunos fenómenos notables que resultan de confinar ondas en una región finita. Esto nos llevará primero a descubrir algunos aspectos particulares relativos a las cuerdas vibrantes, por ejemplo, y luego la generalización de estos hechos nos dará un principio que es probablemente el principio de mayor alcance en la física matemática.

Nuestro primer ejemplo de confinamiento de ondas será limitar una onda por un solo lado. Tomemos el ejemplo simple de una onda unidimensional en una cuerda. Del mismo modo se podría considerar el sonido en una dimensión contra una pared, u otra situación de naturaleza similar, pero el ejemplo de una cuerda será suficiente para los fines que nos hemos propuesto. Supongan que se sujete la cuerda por un extremo, por ejemplo, atándola a una pared "infinitamente sólida". Esto se puede expresar matemáticamente diciendo que en la posición $x = 0$ el desplazamiento y de la cuerda debe ser nulo, porque el extremo no se mueve. Ahora bien, sabemos que si no fuera por la pared la solución general para el movimiento es la suma de dos funciones, $F(x - ct)$ y $G(x + ct)$, la primera representando una onda que viaja en un sentido en la cuerda y la segunda una onda que viaja en el otro sentido en la cuerda:

$$y = F(x - ct) + G(x + ct) \tag{49.1}$$

es la solución general para cualquier cuerda. Pero ahora tenemos que satisfacer la condición de que la cuerda no se mueve en un extremo. Si ponemos $x = 0$ en la ecuación (49.1) y examinamos y para todo valor de t, obtenemos $y = F(-ct) + G(+ct)$. Bien, si esto debe ser cero para todo tiempo, la función $G(ct)$ debe ser $-F(-ct)$. En otras palabras, G de cualquier cosa debe ser $-F$ de menos la misma cosa. Si se introduce este resultado en la ecuación (49.1), encontramos que la solución del problema es

$$y = F(x - ct) - F(-x - ct). \tag{49.2}$$

Es fácil verificar que obtendremos $y = 0$ si ponemos $x = 0$.

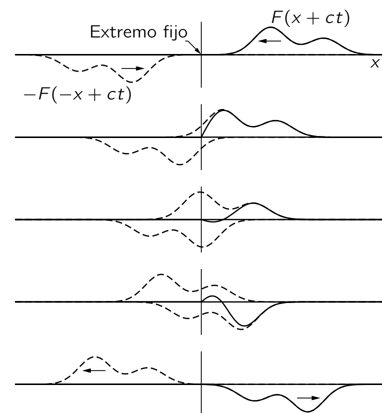


Figura 49-1. Reflexión de una onda como superposición de dos ondas viajeras.

La figura 49-1 muestra una onda que viaja en la dirección x negativa cerca de $x = 0$ y otra onda hipotética que viaja en dirección opuesta con el signo invertido y del otro lado del origen. Decimos hipotética porque, naturalmente, no hay cuerda vibrando de ese lado del origen. Se debe considerar el movimiento total de la cuerda como la suma de estas dos ondas en la región de las x positivas. Cuando llegan al origen, siempre se anularán una a otra en $x = 0$ y finalmente la segunda onda (la reflejada) será la única que existe para x positiva y, naturalmente, estará viajando en dirección opuesta. Estos resultados son equivalentes a la siguiente afirmación: si una onda llega al extremo fijo de una cuerda, se refleja con un cambio de signo. Siempre se puede comprender esa reflexión imaginando que lo que viene del extremo de la cuerda sale invertido desde detrás de la pared. En suma, si suponemos que la cuerda es infinita y que siempre que tenemos una onda que va en un sentido tenemos otra que va en sentido opuesto con la simetría anotada, el desplazamiento en $x = 0$ siempre será nulo y da lo mismo que fijemos o no la cuerda allí.

El siguiente punto a discutir es la reflexión de una onda periódica. Supongan que la onda representada por $F(x - ct)$ es una onda sinusoidal y que se ha reflejado, entonces la onda reflejada $-F(-x - ct)$ también es una onda sinusoidal

de la misma frecuencia, pero que viaja en dirección opuesta. El modo más simple de describir la situación es usando la notación de funciones complejas: $F(x - ct) = e^{i\omega(t-x/c)}$ y $F(-x - ct) = e^{i\omega(t+x/c)}$. Se puede ver que si se sustituye estas expresiones en (49.2) y si se pone x igual 0, resulta $y = 0$ para todo valor de t, por lo que satisface la condición necesaria. Por la propiedad de los exponenciales esto se puede describir en una forma más simple:

$$y = e^{i\omega t}(e^{-i\omega x/c} - e^{i\omega x/c}) = -2ie^{i\omega t}\,\mathrm{sen}\,(\omega x/c). \tag{49.3}$$

Aquí hay algo interesante y nuevo; esta solución nos dice que si observamos en cualquier x fijo, la cuerda oscila con frecuencia ω. ¡Cualquiera que sea este punto, la frecuencia es la misma! Pero hay algunos lugares, en particular cuando $\mathrm{sen}\,(\omega x/c) = 0$, donde no hay ningún desplazamiento. Más aún, si en cualquier instante t tomamos una instantánea de la cuerda vibrante, la fotografía será una onda sinuosidal. Sin embargo, el desplazamiento de esta onda sinusoidal dependerá del tiempo t. Examinando la ecuación (49.3) podemos ver que la longitud de un ciclo de la onda sinusoidal es igual a la longitud de onda de cualquiera de las ondas superpuestas:

$$\lambda = 2\pi c/\omega. \tag{49.4}$$

Los puntos donde no hay movimiento satisfacen la condición $\mathrm{sen}\,(\omega x/c) = 0$, lo cual significa que $(\omega x/c) = 0, \pi, 2\pi, \ldots, n\pi, \ldots$ Estos puntos se llaman *nodos*. Entre dos nodos sucesivos cualesquiera, cada punto se mueve de arriba para abajo sinusoidalmente, pero el diagrama del movimiento permanece fijo en el espacio. Ésta es la característica fundamental de lo que llamamos *modo* de vibración. Si podemos hallar un diagrama de movimiento que tenga la propiedad de que en cualquier punto el objeto se mueva en forma perfectamente sinusoidal y que todos los puntos se muevan con la misma frecuencia (aunque algunos se moverán más que otros), entonces tenemos lo que se llama un modo de vibración.

49-2 Ondas confinadas, con frecuencias naturales

El siguiente problema interesante es considerar lo que ocurre si se sujeta la cuerda por los dos extremos, digamos en $x = 0$ y $x = L$. Podemos comenzar con la idea de la reflexión de las ondas, empezando con algún tipo de protuberancia moviéndose en una dirección. A medida que pasa el tiempo es de esperar que la protuberancia se acerque a un extremo, y cuando pase más tiempo se convertirá en una especie de pequeño bamboleo, porque se combina con la imagen invertida de la protuberancia, que viene del otro lado. Finalmente, la protuberancia original desaparecerá y la protuberancia imagen se moverá en el otro sentido para repetir el proceso en el otro extremo. Este problema tiene una solución fácil, pero una cuestión interesante es si podemos tener un movimiento sinusoidal (la solución que acabamos de describir es *periódica*; pero, por supuesto, no es *sinusoidalmente* periódica). Probemos a poner una onda sinusoidalmente periódica en una cuerda. Si se ata un extremo de la cuerda, sabemos que debe tener el aspecto de nuestra solución anterior (49.3). Si se ata el otro extremo, tiene que tener el mismo aspecto en ese otro extremo. Luego, la única posibilidad de movimiento periódico sinusoidal es que la onda sinusoidal se ajuste perfectamente a la longitud de la cuerda. Si no se ajusta a la longitud de la cuerda, no es una frecuencia natural a la que la cuerda puede seguir oscilando. En resumen, si inicialmente se da a la cuerda una forma de onda sinusoidal que se le ajusta perfectamente, seguirá teniendo esa forma perfecta de onda sinusoidal y oscilará armónicamente con cierta frecuencia.

Matemáticamente, podemos representar la forma por $\mathrm{sen}\, kx$, donde k es igual al factor (ω/c) de las ecuaciones (49.3) y (49.4), y esta función será cero en $x = 0$. Sin embargo, también debe ser cero en el otro extremo. Lo significativo de esto es que k ya no es arbitrario, como en el caso de la cuerda fija en un solo extremo. Con la cuerda fija en ambos extremos, la única posibilidad es que $\mathrm{sen}\,(kL) = 0$, porque ésta es la única condición de que mantendrá ambos extremos fijos. Ahora

bien, para que un seno sea cero, el ángulo debe ser $0, \pi, 2\pi$ o cualquier otro múltiplo entero de π. La ecuación

$$kL = n\pi \tag{49.5}$$

dará, por lo tanto, todos los k posibles, según qué entero se introduzca. Para cada k hay cierta frecuencia ω que, conforme a la ecuación (49.3), es simplemente:

$$\omega = kc = n\pi c/L. \tag{49.6}$$

Hemos encontrado así lo siguiente: una cuerda tiene la propiedad de poder tener movimientos sinusoidales, *pero sólo a ciertas frecuencias*. Ésta es la propiedad más importante de las ondas confinadas. Cualquiera que sea la complicación del sistema, siempre resulta que hay ciertos diagramas de movimiento que tienen una dependencia temporal perfecta, pero con frecuencias que son una propiedad de cada sistema en particular y de la naturaleza de los límites. En el caso de la cuerda tenemos muchas frecuencias posibles diferentes, correspondiendo cada una, por definición, a un modo de vibración, porque un modo es un diagrama de movimiento que se repite sinusoidalmente. La figura 49-2 muestra los tres primeros modos de vibración de una cuerda. La longitud de onda λ del primer modo es $2L$. Esto se puede ver continuando la onda hasta $x = 2L$ para obtener un ciclo completo de la onda sinusoidal. La frecuencia angular ω es $2\pi c$ dividido entre la longitud de onda en general, y en este caso, como λ es $2L$, la frecuencia es $\pi c/L$, lo cual está de acuerdo con (49.6) para $n = 1$. Llamemos ω_1 a la frecuencia del primer modo de vibración. Ahora bien, el modo de vibración siguiente muestra dos lazos con un nodo en medio. Entonces, la longitud de onda de este modo de vibración es simplemente L. El correspondiente valor de k es el doble de grande y la frecuencia es dos veces mayor, es $2\omega_1$. La del tercer modo es $3\omega_1$ y así sucesivamente. Luego, todas las diversas frecuencias de la cuerda son múltiplos, 1, 2, 3, 4, y así sucesivamente, de la frecuencia más baja ω_1.

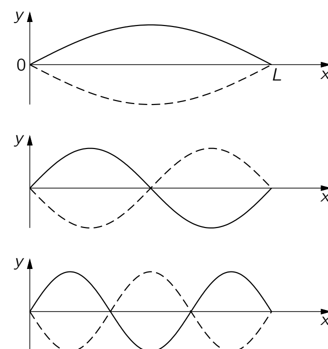


Figura 49-2. Los primeros tres modos de una cuerda vibrante.

Volviendo ahora al movimiento general de la cuerda, resulta que cualquier movimiento posible siempre se puede analizar diciendo que hay más de un modo de vibración al mismo tiempo. En verdad, en un movimiento general un número infinito de modos de vibración debe estar excitado al mismo tiempo. Para tener una idea de esto, mostremos lo que ocurre cuando hay dos modos oscilando al mismo tiempo. Supongan que el primer modo está oscilando como muestra la serie de dibujos de la figura 49.3, que representa la deflexión de la cuerda a intervalos iguales de tiempo durante medio ciclo de la frecuencia más baja.

Ahora bien, suponemos que al mismo tiempo hay una oscilación del segundo modo de vibración. La figura 49-3 también muestra una sucesión de gráficos de este modo de vibración, el cual está, al empezar, desfasado en 90° del primer modo de vibración. Esto significa que al comienzo no tiene desplazamiento, sino que las dos mitades de la cuerda tienen velocidades opuestas. Recordemos ahora un principio general relativo a sistemas lineales: si hay dos soluciones cualesquiera, su suma también es solución. Así, un tercer movimiento posible sería el desplazamiento obtenido sumando las dos soluciones mostradas en la figura 49-3. El resultado, que también se muestra en la figura, comienza a dar la impresión de una protuberancia que corre de un lado a otro entre los extremos de la cuerda, aunque con sólo dos modos de vibración no podemos dar una buena representación, se necesitan más modos. Este resultado es en realidad un caso especial de un gran principio para sistemas lineales:

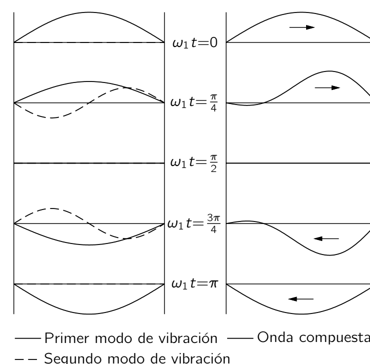


Figura 49-3. Dos modos de vibración se combinan dando una onda viajera.

> *Absolutamente cualquier movimiento se puede analizar suponiendo que es la suma de los movimientos de los modos de vibración diferentes, combinados con amplitudes y fases apropiadas.*

La importancia del principio proviene de que cada modo de vibración es muy simple —es nada más que un movimiento sinusoidal en el tiempo—. Es verdad que aun el movimiento general de una cuerda no es realmente muy complicado, pero hay otros sistemas, la oscilación del ala de un avión, por ejemplo, en los cuales el movimiento es mucho más complicado. No obstante, aun en el ala de

un avión, encontramos que hay cierta manera particular de torcerse que tiene una frecuencia y otras maneras de torcerse que tienen otras frecuencias. Si se puede encontrar estos modos, siempre se puede analizar el movimiento completo como superposición de oscilaciones armónicas (excepto cuando la vibración es de intensidad tal que ya no se puede considerar que el sistema es lineal).

49-3 Modos de vibración en dos dimensiones

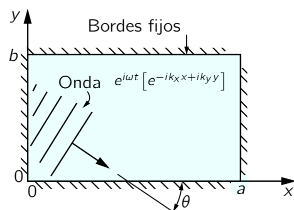


Figura 49-4. Lámina rectangular vibrante.

El siguiente ejemplo a considerar es la interesante situación de modos de vibración en dos dimensiones. Hasta ahora hablamos únicamente de situaciones unidimensionales —una cuerda tensa o las ondas sonoras en un tubo—. Por último tendremos que considerar tres dimensiones, pero un paso más fácil es el de dos dimensiones. Consideren para fijar ideas un parche rectangular de tambor, de goma, limitado de manera que no haya ningún desplazamiento sobre el contorno rectangular, y sean a y b las dimensiones del rectángulo, como muestra la figura 49-4. Ahora preguntamos: ¿cuáles son las características de los movimientos posibles? Podemos empezar con el mismo procedimiento utilizado para la cuerda. Si no hubiese limitación alguna, esperaríamos ondas propagándose con algún movimiento ondulatorio. Por ejemplo, $(e^{i\omega t})(e^{-ik_x x+ik_y y})$ representaría una onda sinusoidal viajando en una dirección que depende de los valores relativos de k_x y k_y. Ahora bien, ¿cómo podemos hacer que el eje x, o sea la línea $y = 0$, sea un nodo? Empleando las ideas desarrolladas para la cuerda unidimensional, podemos imaginar otra onda representada por la función compleja $(-e^{i\omega t})(e^{-ik_x x-ik_y y})$. La superposición de estas ondas dará desplazamiento nulo para $y = 0$ independientemente de los valores de x y de t. (Aunque estas funciones están definidas para y negativa, donde no hay parche de tambor vibrando, esto se puede ignorar ya que el desplazamiento es verdaderamente cero para $y = 0$.) En este caso podemos considerar a la segunda función como onda reflejada.

Sin embargo, queremos una línea nodal en $y = b$ además de que $y = 0$. ¿Cómo hacerlo? La solución está relacionada con algo que hicimos al estudiar la reflexión en cristales. Estas ondas que se cancelan en $y = 0$ harán lo mismo en $y = b$ sólo si $2b \operatorname{sen}\theta$ es un múltiplo entero de λ, donde θ es el ángulo mostrado en la figura 49-4 es:

$$m\lambda = 2b \operatorname{sen}\theta, \qquad m = 0,\, 1,\, 2,\, \ldots \tag{49.7}$$

Ahora podemos hacer en la misma forma que el eje y sea una línea nodal agregando dos funciones más, $-(e^{i\omega t})(e^{+ik_x x+ik_y y})$ y $+(e^{i\omega t})(e^{+ik_x x-ik_y y})$, cada una de las cuales representa una reflexión de una de las otras dos ondas que hay en la línea $x = 0$. La condición de que haya una línea nodal en $x = a$ es similar a la de $y = b$. Es que $2a\cos\theta$ también debe ser un múltiplo entero de λ:

$$n\lambda = 2a\cos\theta. \tag{49.8}$$

El resultado final es entonces que las ondas que rebotan de un lado a otro en la caja producen un diagrama de ondas estacionarias, o sea un modo de vibración definido.

Así, pues, debemos satisfacer las dos condiciones anteriores si queremos tener un modo de vibración. Encontremos primero la longitud de onda. Se puede obtener eliminando el ángulo θ con (49.7) y (49.8), obteniendo la longitud de onda en función de a, b, n y m. La manera más fácil de hacerlo es dividir ambos miembros de las ecuaciones entre $2b$ y $2a$, respectivamente, elevar las ecuaciones al cuadrado y sumarlas. El resultado es $\operatorname{sen}^2\theta + \cos^2\theta = 1 = (n\lambda/2a)^2 + (m\lambda/2b)^2$, de donde se puede despejar λ:

$$\frac{1}{\lambda^2} = \frac{n^2}{4a^2} + \frac{m^2}{4b^2}. \tag{49.9}$$

De esta manera hemos determinado la longitud de onda en función de dos enteros, y de la longitud de onda obtenemos inmediatamente la frecuencia ω porque, como sabemos, la frecuencia es igual a $2\pi c$ dividido entre la longitud de onda.

Este resultado es tan interesante e importante que deberíamos deducirlo con un análisis puramente matemático en vez de hacerlo razonando sobre las reflexiones. Representemos la vibración con una superposición de cuatro ondas elegidas de manera que las cuatro líneas $x = 0$, $x = a$, $y = 0$ y $y = b$ sean nodos. Además, haremos que todas las ondas tengan la misma frecuencia de manera que el movimiento resultante represente un modo de vibración. Conforme al tratamiento que dimos anteriormente de la reflexión de la luz, sabemos que $(e^{i\omega t})(e^{-ik_x x+ik_y y})$ representa una onda que se propaga en la dirección indicada en la figura 49-4. La ecuación (49.6), es decir $k = \omega/c$, es válida siempre que:

$$k^2 = k_x^2 + k_y^2. \tag{49.10}$$

Está claro en la figura que $k_x = k\cos\theta$ y $k_y = k\,\mathrm{sen}\,\theta$.

Ahora nuestra ecuación para el desplazamiento, llamemos ϕ, del parche rectangular de tambor adopta la gran forma:

$$\phi = [e^{i\omega t}][e^{(-ik_x x+ik_y y)} - e^{(+ik_x x+ik_y y)} - e^{(-ik_x x-ik_y y)} + e^{(+ik_x x-ik_y y)}]. \tag{49.11a}$$

Aunque esto parece más bien un lío, la suma de estas cosas no es tan difícil. Se pueden combinar las exponenciales para dar funciones seno, de modo que el desplazamiento resulta:

$$\phi = [4\,\mathrm{sen}\,k_x x\,\mathrm{sen}\,k_y y][e^{i\omega t}]. \tag{49.11b}$$

En otras palabras, es una oscilación sinusoidal, sí, con un diagrama que también es sinusoidal tanto en la dirección x como en la y. Por supuesto que nuestras condiciones de contorno están satisfechas en $x = 0$ y en $y = 0$. También queremos que ϕ sea cero para $x = a$ y para $y = b$. En consecuencia, tenemos que introducir otras dos condiciones, $k_x a$ debe ser un múltiplo entero de π y $k_y b$ debe ser otro múltiplo entero de π. Como hemos visto que $k_x = k\cos\theta$ y $k_y = k\,\mathrm{sen}\,\theta$, obtenemos inmediatamente las ecuaciones (49.7) y (49.8) y de ellas el resultado final (49.9).

Tomemos ahora como ejemplo un rectángulo cuyo largo es dos veces su alto. Si tomamos $a = 2b$ y utilizamos las ecuaciones (49.4) y (49.9), podemos calcular la frecuencia de todos los modos de vibración:

Tabla 49-1

Forma del modo de vibración	m	n	$(\omega/\omega_0)^2$	ω/ω_0
+	1	1	1.25	1.12
+ ¦ −	1	2	2.00	1.41
+ ¦ − ¦ +	1	3	3.25	1.80
− / +	2	1	4.25	2.06
− ¦ + / + ¦ −	2	2	5.00	2.24

$$\omega^2 = \left(\frac{\pi c}{b}\right)^2 \frac{4m^2 + n^2}{4}. \tag{49.12}$$

La tabla 49-1 da algunos de los modos de vibración simples y muestra también su forma de manera cualitativa.

El punto más importante a recalcar sobre este caso particular es que las frecuencias no son múltiplos unas de las otras, ni son múltiplos de ningún número. La idea de que las frecuencias naturales están relacionadas armónicamente no es cierta en general. No es válido para un sistema de más de una dimensión ni es verdadera para sistemas unidimensionales que sean más complicados que una cuerda con densidad y tensión uniformes. Un ejemplo de esto último es una cadena colgante en la cual la tensión es mayor arriba que abajo. Si se pone esa cadena a oscilar armónicamente, hay varias frecuencias y modos de vibración, pero ni las frecuencias son múltiplos simples de algún número, ni la forma de los modos de vibración es sinusoidal.

Los modos de vibración de sistemas más complicados son aún más complejos. Por ejemplo, dentro de la boca tenemos una cavidad encima de las cuerdas vocales, y moviendo la lengua y los labios, etc., hacemos un tubo con el extremo abierto o uno con el extremo cerrado de diferentes diámetros y formas; es un resonador, terriblemente complicado, pero no obstante es un resonador. Ahora, cuando se habla, las cuerdas vocales producen cierto tipo de tono. El tono es bastante complicado y hay muchos sonidos que salen, pero la cavidad de la boca modifica aún más ese tono debido a las diversas frecuencias resonantes de la cavidad. Por ejemplo, un cantante puede entonar varias vocales como a, o, u, etc., todas a la misma altura, pero suenan diferentes porque en esta cavidad los diversos armónicos están en resonancia en grados diferentes. La grandísima importancia de las frecuencias resonantes de una cavidad en la modificación de los sonidos de la voz se puede demostrar por medio de un experimento simple. Como la velocidad del sonido varía con la inversa de la raíz cuadrada de la densidad, se puede variar la velocidad del sonido usando gases diferentes. Si se usa helio en vez de aire, de modo que la densidad es menor, la velocidad del sonido es mucho mayor y todas las frecuencias de una cavidad se elevarán. Así, si uno se llena los pulmones de helio antes de hablar, las características de la voz se alterarán totalmente aun cuando las cuerdas vocales estén vibrando a la misma frecuencia.

49-4 Péndulos acoplados

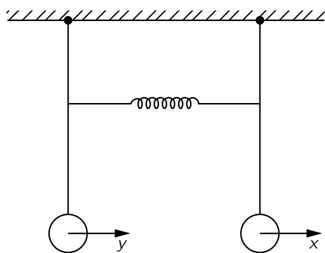


Figura 49-5. Dos péndulos acoplados.

Finalmente, debemos recalcar que no solamente existen modos de vibración en sistemas continuos complicados, sino también en sistemas mecánicos muy simples. Un buen ejemplo es el sistema de dos péndulos acoplados discutido en el capítulo precedente. En ese capítulo se demostró que se podía analizar el movimiento como superposición de dos movimientos armónicos de frecuencia diferente. Así, pues, hasta este sistema se puede analizar en términos de movimientos armónicos o modos. La cuerda tiene un número infinito de modos y la superficie bidimensional también tiene un número infinito de modos. En cierto sentido es una doble infinidad, si es que sabemos cómo contar infinitos. Pero un simple objeto mecánico que tiene únicamente dos grados de libertad y que sólo necesita dos variables para ser descrito, tiene solamente dos modos.

Hagamos un análisis matemático de estos dos modos para el caso en que los péndulos tengan la misma longitud. Sea x el desplazamiento de uno y y el del otro, como muestra la figura 49-5. Sin el resorte, la fuerza sobre la primera masa es proporcional al desplazamiento de la misma a causa de la gravedad. Si no hubiera resorte, habría cierta frecuencia natural ω_0 para este péndulo solo. La ecuación de movimiento sin el resorte sería:

$$m\,\frac{d^2x}{dt^2} = -m\omega_0^2 x. \tag{49.13}$$

El otro péndulo oscilaría de la misma manera si no hubiera resorte. Además de la fuerza restauradora debida a la gravedad, también hay una fuerza adicional

que tira de la primera masa. Esa fuerza depende de la distancia en exceso de x respecto a y y es proporcional a esa diferencia, por lo que es cierta constante que depende de la geometría multiplicada por $(x - y)$. La misma fuerza en sentido inverso actúa sobre la segunda masa. En consecuencia, las ecuaciones de movimiento a resolver son:

$$m\,\frac{d^2x}{dt^2} = -m\omega_0^2 x - k(x - y), \qquad m\,\frac{d^2y}{dt^2} = -m\omega_0^2 y - k(y - x). \tag{49.14}$$

Para hallar un movimiento en que las dos masas se mueven a la misma frecuencia, debemos determinar cuánto se mueve cada masa. En otras palabras, el péndulo x y el péndulo y oscilarán con la misma frecuencia, pero sus amplitudes deben tener ciertos valores, A y B, cuya relación es fija. Probemos esta solución:

$$x = Ae^{i\omega t}, \qquad y = Be^{i\omega t}. \tag{49.15}$$

Si se sustituyen en la ecuación (49.14) y se juntan los términos análogos, los resultados son:

$$\begin{aligned}\left(\omega^2 - \omega_0^2 - \frac{k}{m}\right)A &= -\frac{k}{m}\,B,\\ \left(\omega^2 - \omega_0^2 - \frac{k}{m}\right)B &= -\frac{k}{m}\,A.\end{aligned} \tag{49.16}$$

Para obtener las ecuaciones tal como están escritas, se ha eliminado el factor común $e^{i\omega t}$ y se las ha dividido entre m.

Ahora bien, vemos que tenemos dos ecuaciones para lo que aparentemente son dos incógnitas. Pero en realidad no hay *dos* incógnitas porque la amplitud total del movimiento es algo que no podemos determinar a partir de estas ecuaciones. Las ecuaciones precedentes sólo pueden determinar el *cociente* entre A y B, *pero ambas deben dar el mismo cociente*. La necesidad de que estas dos ecuaciones sean compatibles requiere que la frecuencia sea algo muy especial.

En este caso particular, esto se puede hacer bastante fácil si se multiplica una ecuación por otra, el resultado será:

$$\left(\omega^2 - \omega_0^2 - \frac{k}{m}\right)^2 AB = \left(\frac{k}{m}\right)^2 AB. \tag{49.17}$$

El factor AB se puede eliminar de ambos miembros a no ser que A y B sean cero, lo cual significa que no hay ningún movimiento. Si hay movimiento, los otros factores deben ser iguales dando una ecuación cuadrática a resolver. El resultado es que hay dos frecuencias posibles:

$$\omega_1^2 = \omega_0^2, \qquad \omega_2^2 = \omega_0^2 + \frac{2k}{m}. \tag{49.18}$$

Más aún, si se vuelven a sustituir estos valores de la frecuencia en la ecuación (49.16), encontramos que para la primera frecuencia $A = B$ y para la segunda, $A = -B$. Éstas son las “formas de los modos de vibración”, como se puede verificar fácilmente con el experimento.

Está claro que en el primer modo, para el cual $A = B$, el resorte nunca se estira y ambas masas oscilan con frecuencia ω_0 como si el resorte estuviera ausente. En la otra solución, para la cual $A = -B$, el resorte contribuye con una fuerza restauradora y aumenta la frecuencia. Resulta un caso más interesante si los dos péndulos tienen longitudes diferentes. El análisis es muy parecido al dado más arriba y se deja al lector como ejercicio.

49-5 Sistemas lineales

Resumamos ahora las ideas discutidas anteriormente, que son todos los aspectos de lo que posiblemente sea el principio más general y maravilloso de la

física matemática. Si tenemos un sistema lineal cuyas características son independientes del tiempo, el movimiento no tiene por qué tener ninguna simplicidad particular; en realidad puede ser extremadamente complejo, pero hay movimientos muy especiales; es común una serie de movimientos especiales, en los cuales el diagrama global del movimiento varía exponencialmente con el tiempo. En los sistemas vibrantes de que hemos estado hablando, la exponencial es imaginaria y en vez de decir "exponencialmente", preferimos decir "sinusoidalmente" con el tiempo. Sin embargo, se puede ser más general y decir que los movimientos variarán exponencialmente con el tiempo en modos muy especiales, con formas muy especiales. Siempre se puede representar el movimiento más general de un sistema como una superposición de movimientos en los que interviene cada una de las diversas exponenciales.

Vale la pena reafirmar esto para el caso de un movimiento sinusoidal. Un sistema lineal no tiene por qué moverse con un movimiento puramente sinusoidal, es decir, con una sola frecuencia definida, pero cualquiera que sea la forma en que se mueva, se puede representar este movimiento como superposición de movimientos puramente sinusoidales. La frecuencia de cada uno de estos movimientos es una característica del sistema, y el perfil, o forma de onda, de cada uno de estos movimientos que también es una característica del sistema. El movimiento general de cualquiera de estos sistemas se puede caracterizar dando la intensidad y la fase de cada uno de estos modos y sumándolos todos. Otra manera de decir esto es que cualquier sistema vibrante lineal es equivalente a un conjunto de osciladores armónicos independientes, con las frecuencias naturales correspondientes a los modos de vibración.

Concluimos este capítulo haciendo notar la conexión entre modos de vibración y mecánica cuántica. En la mecánica cuántica lo que vibra, o lo que varía en el espacio, es la amplitud de una función de probabilidad que da la probabilidad de encontrar un electrón, o un sistema de electrones, en una configuración determinada. Esta función amplitud puede variar en el espacio y en el tiempo y satisface, de hecho, una ecuación lineal. Pero en la mecánica cuántica hay una transformación; lo que llamamos frecuencia de la amplitud de probabilidad, que es igual, en el concepto clásico, a energía. En consecuencia, podemos traducir a este caso el principio enunciado más arriba tomando la palabra *frecuencia* y remplazándola por *energía*. Se convierte en algo así como: no es necesario que un sistema cuántico, un átomo por ejemplo, tenga una energía definida, como no es necesario que un sistema mecánico simple tenga una frecuencia definida; pero, sea cualesquiera la manera en que se comporta el sistema, siempre se puede representar su comportamiento como una superposición de estados de energía definida. La energía de cada estado es una característica del átomo y lo mismo es el diagrama de amplitud que determina la probabilidad de encontrar partículas en diversos lugares. El movimiento general se puede describir dando la amplitud de cada uno de estos estados diferentes de energía. En la mecánica cuántica éste es el origen de los niveles de energía. Como la mecánica cuántica está representada por ondas, cuando el electrón no tiene energía suficiente para escapar a la larga del protón, ésas son *ondas confinadas*. Como en las ondas confinadas de una cuerda, hay frecuencias definidas para la solución de la ecuación de onda para la mecánica cuántica. La interpretación cuántica es que éstas son *energías* definidas. En consecuencia, un sistema cuántico puede tener estados definidos de energía fija porque está representado por ondas; ejemplos de esto son los niveles de energía de los diversos átomos.

50

Armónicos

50-1 Tonos musicales

Se dice que Pitágoras descubrió que cuando se hace sonar a la vez dos cuerdas similares bajo la misma tensión y que difieren sólo en longitud, dan un efecto que es agradable al oído *si* las longitudes de las cuerdas están en proporción de dos números enteros pequeños. Si las longitudes son como uno es a dos, corresponden entonces a la octava en música. Si las longitudes son como dos es a tres, corresponden al intervalo entre *do* y *sol*, que se llama una quinta. Estos intervalos se aceptan generalmente como acordes que suenan "agradable".

Pitágoras se impresionó tanto con este descubrimiento que lo hizo la base de una escuela —se llamaron pitagóricos— que tuvo creencias místicas en los grandes poderes de los números. Se pensaba que algo parecido se encontraría en los planetas —o "esferas"—. Algunas veces oímos la expresión: "la música de las esferas". La idea fue que habría algunas relaciones numéricas entre las órbitas de los planetas o entre otras cosas en la naturaleza. La gente cree generalmente que esto fue solamente una especie de superstición mantenida por los griegos. Pero ¿es esto tan diferente de nuestro interés científico en las relaciones cuantitativas? El descubrimiento de Pitágoras fue el primer ejemplo, fuera de la geometría, de una relación numérica en la naturaleza. Debe haber sido muy sorprendente el descubrir de repente que había un *hecho* de la naturaleza que involucraba una relación numérica sencilla. Simples medidas de longitudes dieron una predicción de algo que no tenía conexión aparente con la geometría —la producción de sonidos agradables—. Este descubrimiento condujo a la extensión de que quizás una buena herramienta para la comprensión de la naturaleza sería el análisis aritmético y matemático. Los resultados de la ciencia moderna justifican este punto de vista.

Pitágoras solamente pudo haber hecho su descubrimiento mediante una observación experimental. No obstante, este aspecto importante parece no haberle impresionado. De lo contrario, la física hubiera tenido un comienzo más temprano. (¡Siempre es fácil reconsiderar lo que alguien ha hecho y decidir lo que él *debería* haber hecho!)

Podemos observar un tercer aspecto de este interesante descubrimiento. El descubrimiento tiene que ver con dos notas que *suenan agradable* al oído. Podemos preguntarnos si *nosotros* hemos avanzado más que Pitágoras en la comprensión de *por qué* sólo ciertos sonidos son agradables al oído. La teoría general de la estética no está probablemente más adelantada ahora que en el tiempo de Pitágoras. En este descubrimiento de los griegos encontramos los tres aspectos: experimento, relaciones matemáticas y estética. La física ha tenido un gran adelanto solamente en las dos primeras partes. Este capítulo tratará sobre nuestra comprensión actual del descubrimiento de Pitágoras.

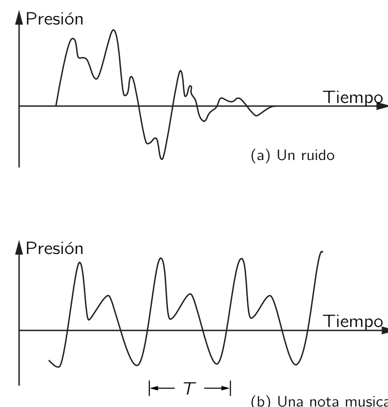


Figura 50-1. Presión en función del tiempo para (a) un ruido y (b) una nota musical.

Entre los sonidos que oímos hay una clase que llamamos *ruidos*. El ruido corresponde a una especie de vibración irregular del tímpano producida por una vibración irregular de algún objeto cercano. Si hacemos un diagrama para indicar la presión del aire en el tímpano (y, por lo tanto, el desplazamiento del mismo) en función del tiempo, la gráfica que corresponde a un ruido puede parecerse a la que representa la figura 50-1(a). (Tal ruido podría corresponder aproximadamente al sonido de un zapatazo.) El sonido de la *música* tiene un carácter diferente. La música se caracteriza por la presencia de *tonos* más o menos *sostenidos* —o "notas" musicales—. (¡Los instrumentos musicales también pueden hacer ruidos!) El tono puede durar un tiempo relativamente corto, como cuando se presiona

una tecla en un piano, o se puede prolongar casi indefinidamente, como cuando un flautista mantiene una nota larga.

¿Cuál es el carácter especial de una nota musical desde el punto de vista de la presión en el aire? Una nota musical difiere de un ruido en que hay una periodicidad en su gráfica. Hay una cierta forma irregular de la variación de presión del aire con el tiempo y la forma se repite una y otra vez. La figura 50-1(b) muestra un ejemplo de la función presión respecto al tiempo que correspondería a una nota musical.

Los músicos hablan corrientemente de un tono musical en términos de tres características: intensidad, tono y "timbre". La "intensidad" corresponde a la magnitud de los cambios de presión. El "tono" corresponde al periodo de tiempo para una repetición de la función básica de presión. (Las notas "bajas" tienen periodos más grandes que las notas "altas".) El "timbre" de una nota tiene que ver con las diferencias que somos capaces de oír entre dos notas de la misma intensidad y tono. Un oboe, un violín y una soprano se pueden distinguir aun cuando den notas del mismo tono. El timbre tiene que ver con la estructura del diagrama que se repite.

Consideremos por un momento el sonido producido por una cuerda vibrante. Si accionamos la cuerda, tirando de ella y soltándola, el movimiento subsiguiente estará determinado por los movimientos de las ondas que hemos producido. Sabemos que estas ondas viajarán en ambas direcciones y que se reflejarán en los extremos. Irán hacia adelante y hacia atrás durante mucho tiempo. Aunque la onda sea muy complicada, se repetirá. El periodo de repetición es justamente el tiempo T necesario para que la onda viaje dos longitudes completas de la cuerda. Porque éste es exactamente el tiempo que necesita cualquier onda, una vez que ha comenzado, para reflejarse en cada extremo, volver a su posición inicial y seguir en la dirección original. El tiempo es el mismo para ondas que comiencen en cualquier dirección. Cada punto de la cuerda volverá, pues, a su posición inicial después de un periodo, y de nuevo un periodo más tarde, etc. La onda de sonido producida debe tener también la misma repetición. Vemos por qué una cuerda accionada produce una nota musical.

50-2 La serie de Fourier

Discutimos en el capítulo precedente otro modo de considerar el movimiento de un sistema vibrante. Hemos visto que una cuerda tiene diversos modos naturales de oscilación y que cualquier clase particular de vibración que se pueda originar por las condiciones originales se puede considerar como una combinación —en proporciones convenientes— de varios de los modos naturales oscilando a la vez. Encontramos que los modos normales de oscilación para una cuerda tenían las frecuencias $\omega_0, 2\omega_0, 3\omega_0, \ldots$ El movimiento más general de una cuerda que ha sido pulsada, por lo tanto, está compuesto de la suma de una oscilación sinusoidal a la frecuencia fundamental ω_0, otra a la frecuencia del segundo armónico $2\omega_0$, otra a la del tercer armónico $3\omega_0$ etc. Ahora bien, el modo fundamental se repite cada periodo $T_1 = 2\pi/\omega_0$. El segundo armónico se repite cada $T_2 = 2\pi/2\omega_0$. *También* se repite cada $T_1 = 2T_2$, después de *dos* de sus periodos. Análogamente, el tercer modo armónico se repite después de un tiempo T_1 equivalente a 3 de sus periodos. Vemos de nuevo por qué una cuerda pulsada repite todo su diagrama con periodicidad T_1. Produce una nota musical.

Hemos estado hablando del movimiento de una cuerda. Pero el *sonido*, que es el movimiento del aire, es producido por el movimiento de la cuerda, por lo que sus vibraciones también deben estar compuestas de los mismos armónicos —aunque no estamos pensando ya en los modos normales del aire—. Igualmente, la intensidad relativa de los armónicos puede ser diferente en el aire que en la cuerda, particularmente si la cuerda está "acoplada" al aire mediante una caja de resonancia. La eficiencia del acoplamiento al aire es diferente para distintos armónicos.

Si dejamos que $f(t)$ represente la presión del aire en función del tiempo para una nota musical, como la de la figura 50-1(b), entonces esperamos que se pueda

escribir $f(t)$ como suma de un cierto número de funciones armónicas simples del tiempo —como $\cos\omega t$— para cada una de las diversas frecuencias armónicas. Si el periodo de la vibración es T, la frecuencia angular fundamental será $\omega = 2\pi/T$ y los armónicos serán $2\omega, 3\omega$, etcétera.

Hay una pequeña complicación. Para cada frecuencia podemos esperar que las fases iniciales no serán necesariamente las mismas para todas las frecuencias. Por lo tanto, deberíamos usar funciones como $\cos(\omega t + \phi)$. Sin embargo, es más sencillo usar en su lugar funciones seno y coseno para cada frecuencia. Recordemos que:

$$\cos(\omega t + \phi) = (\cos\phi\cos\omega t - \operatorname{sen}\phi\operatorname{sen}\omega t) \tag{50.1}$$

y como ϕ es una constante, *cualquier* oscilación sinusoidal de frecuencia ω se puede escribir como suma de un término con $\cos\omega t$ y otro término con $\operatorname{sen}\omega t$.

Concluimos, entonces, que *cualquier* función $f(t)$ periódica con periodo T se puede escribir matemáticamente en la forma:

$$\begin{aligned} f(t) = a_0& \\ &+ a_1\cos\ \ \omega t + b_1\operatorname{sen}\ \ \omega t \\ &+ a_2\cos 2\omega t + b_2\operatorname{sen} 2\omega t \\ &+ a_3\cos 3\omega t + b_3\operatorname{sen} 3\omega t \\ &+ \cdots \qquad\quad + \cdots \end{aligned} \tag{50.2}$$

donde $\omega = 2\pi/T$ y las a y b son constantes numéricas que nos dicen cuánto de cada oscilación componente está presente en la oscilación $f(t)$. Hemos puesto el término a_0 de "frecuencia cero" para que nuestra fórmula sea completamente general, aunque generalmente es cero para una nota musical. Representa una desviación del valor promedio (es decir, el nivel "cero") de la presión de sonido. Con él nuestra fórmula se puede aplicar a cualquier caso. La igualdad de la ecuación (50.2) está representada esquemáticamente en la figura 50-2. (Se deben escoger convenientemente las amplitudes a_n y b_n de las funciones armónicas. Están mostradas esquemáticamente y sin una escala particular en la figura.) La serie (50.2) se llama *serie de Fourier* de $f(t)$.

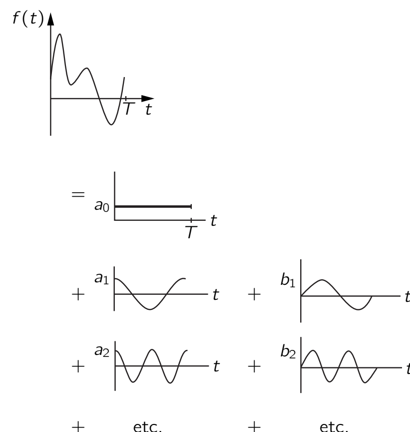


Figura 50-2. Cualquier función periódica $f(t)$ es igual a una suma de funciones armónicas simples.

Hemos dicho que *cualquier* función periódica se puede construir de este modo. Deberíamos corregir eso y decir que cualquier onda de sonido, o cualquier función de las que ordinariamente encontramos en física se puede construir mediante tal suma. Los matemáticos pueden inventar funciones que no se pueden construir a partir de funciones armónicas simples —por ejemplo, una función que tiene una "vuelta hacia atrás"—, ¡de modo que tiene dos valores para algunos valores de t! No tenemos por qué preocuparnos de esas funciones aquí.

50-3 Timbre y consonancia

Ahora estamos en condiciones de describir qué es lo que determina el "timbre" de una nota musical. Es la cantidad relativa de los diversos armónicos —los valores de las a y b—. Una nota con el primer armónico solamente es una nota "pura". Una nota con muchos armónicos fuertes es una nota "rica". Un violín produce una proporción de armónicos diferente de la que produce un oboe.

Podemos "fabricar" diversas notas musicales si conectamos diversos "osciladores" a un altoparlante. (Un oscilador produce frecuentemente una función armónica simple casi pura.) Deberíamos escoger las frecuencias de los osciladores de manera que sean $\omega, 2\omega, 3\omega$, etc. Ajustando el control de volumen de cada oscilador, podemos añadir cualquier cantidad que deseemos de cada armónico —y por consiguiente producir notas de diferente timbre—. Un órgano eléctrico trabaja de un modo parecido. Las "teclas" seleccionan la frecuencia del oscilador fundamental y las "clavijas" son llaves que controlan la proporción relativa de los armónicos. Usando estas llaves se puede hacer que el órgano suene como una flauta, un oboe o un violín.

Es interesante que para producir tales notas "artificiales" necesitamos solamente un oscilador para cada frecuencia —no necesitamos osciladores separados para las componentes seno y coseno—. El oído no es muy sensible a las fases relativas de los armónicos. Presta atención principalmente al *total* de las partes seno y coseno de cada frecuencia. Nuestro análisis es más exacto de lo necesario para explicar el aspecto *subjetivo* de la música. Sin embargo, la respuesta de un micrófono o de cualquier otro instrumento físico sí depende de las fases y nuestro análisis completo se puede necesitar para tratar tales casos.

El "timbre" de un sonido hablado también determina los sonidos de vocales que reconocemos en el lenguaje. La forma de la boca determina las frecuencias de los modos naturales de vibración del aire en la boca. Algunos de estos modos se ponen en vibración mediante las ondas de sonido provenientes de las cuerdas vocales. De manera que las amplitudes de algunos de los armónicos del sonido se aumentan respecto a otras. Cuando cambiamos la forma de nuestra boca, damos preferencia a armónicos de frecuencias diferentes. Estos efectos cuentan para la diferencia entre un sonido "e e e" y un sonido "a a a".

Todos sabemos que un sonido vocal particular —"e e e", digamos— aún "suena como" la misma vocal, ya lo digamos (o cantemos) en un tono alto o bajo. Del mecanismo que describimos, esperaríamos que se acentuasen frecuencias *particulares* cuando colocamos nuestra boca para una "e e e" y que *no* cambiasen cuando nosotros cambiamos el tono de nuestra voz. Así, la relación de los armónicos importantes al fundamental —esto es, el "timbre"— cambia cuando nosotros cambiamos el tono. Aparentemente el mecanismo mediante el cual reconocemos el lenguaje no se basa en relaciones armónicas específicas.

¿Qué diríamos ahora acerca del descubrimiento de Pitágoras? Comprendemos que dos cuerdas semejantes con longitudes en la proporción de 2 a 3 tendrán frecuencias fundamentales en la proporción de 3 a 2. Pero ¿por qué deberían "sonar agradable" juntas? Quizá deberíamos buscar la explicación en las frecuencias de los armónicos. El segundo armónico más bajo de la cuerda más corta tendrá *igual* frecuencia que el tercer armónico de la cuerda más larga. (Es fácil demostrar —o creer que— una cuerda pulsada produce con mucha intensidad los diversos armónicos más bajos.)

Quizás deberíamos dar las siguientes reglas. Las notas son consonantes cuando tienen armónicos de la misma frecuencia. Las notas son disonantes si sus armónicos superiores tienen frecuencias cercanas, pero lo bastante separadas para que haya pulsaciones rápidas entre las dos. Por qué las pulsaciones no suenan agradables y por qué los unísonos de los armónicos superiores suenan agradables es algo que no sabemos definir o describir. No podemos decir a partir de este conocimiento de lo que *suena* bien, lo que debería, por ejemplo, *oler* bien. En otras palabras, nuestro conocimiento de ello no es algo más general que el aserto de que cuando están al unísono suenan bien. No nos permite deducir nada más que las propiedades de la armonía en música.

Es fácil comprobar las relaciones armónicas que hemos descrito mediante algunos experimentos sencillos con el piano. Llamemos a los tres do sucesivos por la mitad del teclado $\mathrm{do}_1, \mathrm{do}_2, \mathrm{do}_3$, y a los sol inmediatamente superiores, $\mathrm{sol}_1, \mathrm{sol}_2, \mathrm{sol}_3$. Entonces tendremos las frecuencias relativas fundamentales como:

$$\begin{array}{ll} \mathrm{do}_1 - 2 & \mathrm{sol}_1 - 3 \\ \mathrm{do}_2 - 4 & \mathrm{sol}_2 - 6 \\ \mathrm{do}_3 - 8 & \mathrm{sol}_3 - 12 \end{array}$$

Estas relaciones armónicas se pueden demostrar en la siguiente forma: supongan que presionamos do_2 *despacio* —de esa manera no suena, pero hacemos que el amortiguador se levante—. Si entonces hacemos sonar do_1 producirá su fundamental propio *y* algunos armónicos secundarios. El segundo armónico pondrá en vibración las cuerdas de do_2. Si soltamos do_1 (presionando aún do_2) el amortiguador detendrá la vibración de las cuerdas de do_1, y podemos oír (suavemente) la nota do_2 que se va apagando. De un modo semejante, el tercer armónico de do_1 puede causar una vibración de sol_2. O bien, el sexto

de do_1 (haciéndose ahora mucho más débil) puede ocasionar una vibración del fundamental de sol_3.

Se obtiene un resultado algo diferente si presionamos sol_1 con cuidado y hacemos sonar do_2. El tercer armónico de do_2 corresponderá al cuarto armónico de sol_1, por lo que *solamente* se excitará el cuarto armónico de sol_1. Podemos oír (si escuchamos de cerca) el sonido de sol_3 ¡que está dos octavas por encima del sol_1 que hemos apretado! Es fácil imaginar otras combinaciones para este juego.

Podemos señalar de paso que la escala mayor se puede definir justamente mediante la condición de que los tres acordes mayores (fa-la-do), (do-mi-sol) y (sol-si-re) representen *cada* uno secuencias de notas con la relación de frecuencias $(4 : 5 : 6)$. Estas relaciones —además del hecho de que una octava (do_1 - do_2, si_1 - si_2, etc.) tiene la relación $1 : 2$— determinan la escala total para el caso "ideal", o sea para lo que se llama "entonación justa". Los instrumentos de teclado como el piano *no* se afinan usualmente así, pero se hace un poco de "fraude" para que las frecuencias sean *aproximadamente* correctas para todos los posibles tonos de partida. Para esta afinación, que se llama "templado", la octava (aun $1 : 2$) se divide en 12 intervalos iguales para los cuales la relación de frecuencia es $(2)^{1/12}$. Una quinta ya no tiene la relación de frecuencia $3/2$, sino $2^{7/12} = 1.499$, que aparentemente es bastante aproximada para la mayoría de los oídos.

Hemos establecido una regla de consonancia en términos de la coincidencia de armónicos. ¿Es quizás esta coincidencia la *razón* de que dos notas sean consonantes? Un investigador ha sostenido que dos notas *puras* —notas fabricadas cuidadosamente para estar libres de armónicos— no dan las *sensaciones* de consonancia o disonancia cuando las frecuencias relativas están colocadas en o cerca de las relaciones esperadas. (Tales experimentos son difíciles porque es difícil fabricar notas puras por motivos que más adelante veremos.) Aún no podemos estar seguros de si el oído está apareando armónicos o haciendo aritmética cuando nosotros decidimos que nos agrada un sonido.

50-4 Coeficientes de Fourier

Volvamos ahora a la idea de que cualquier nota —esto es, un sonido *periódico*— se puede representar mediante una combinación apropiada de armónicos. Queríamos demostrar cómo podemos encontrar qué cantidad se requiere de cada armónico. Es fácil, naturalmente, calcular $f(t)$ usando la ecuación (50.2), si se *nos dan* todos los coeficientes a y b. La cuestión es ahora, si se nos da $f(t)$, ¿cómo podemos conocer cuáles deberían ser los coeficientes de los diversos términos armónicos? (Es fácil hacer una torta siguiendo una receta; pero, ¿podemos escribir la receta si se nos da la torta?)

Fourier descubrió que en realidad no era muy difícil. El término a_0 es fácil. Acabamos de decir que es justamente el valor promedio de $f(t)$ en un periodo (desde $t = 0$, hasta $t = T$). Podemos ver que esto es realmente así. El valor medio de una función seno o coseno sobre un periodo es cero. Sobre dos, o tres, o cualquier número entero de periodos, también es cero. Por lo que el valor medio de todos los términos del segundo miembro de la ecuación (50.2) es cero, excepto para a_0. (Recuerden que debemos tomar $\omega = 2\pi/T$.)

Ahora bien, el promedio de una suma es la suma de los promedios. Por lo que el promedio de $f(t)$ es solamente el promedio de a_0. Pero a_0 es una *constante*, por lo que su promedio es justamente su valor. Recordando la definición de promedio, tenemos:

$$a_0 = \frac{1}{T}\int_0^T f(t)\,dt. \tag{50.3}$$

Los otros coeficientes son solamente un poquito más difíciles. Para encontrarlos podemos usar un truco descubierto por Fourier. Supongan que multiplicamos los dos miembros de la ecuación (50.2) por alguna función armónica —por $\cos 7\omega t$

digamos—. Tenemos, entonces

$$\begin{aligned} f(t)\cdot\cos 7\omega t = a_0\cdot\cos 7\omega t \\ &+ a_1\cos\ \ \omega t\cdot\cos 7\omega t + b_1\,\text{sen}\ \ \omega t\cdot\cos 7\omega t \\ &+ a_2\cos 2\omega t\cdot\cos 7\omega t + b_2\,\text{sen}\,2\omega t\cdot\cos 7\omega t \\ &+\cdots \qquad\qquad\qquad\quad +\cdots \\ &+ a_7\cos 7\omega t\cdot\cos 7\omega t + b_7\,\text{sen}\,7\omega t\cdot\cos 7\omega t \\ &+\cdots \qquad\qquad\qquad\quad +\cdots \end{aligned} \tag{50.4}$$

Ahora, promediemos ambos miembros. El promedio de $a_0\cos 7\omega t$ sobre el tiempo T es proporcional al promedio de un coseno sobre 7 periodos enteros. Pero esto es justamente cero. El promedio de *casi todos* los términos restantes *también* es cero. Consideremos el término a_1. Sabemos, en general, que:

$$\cos A\cos B = \tfrac{1}{2}\cos{(A+B)} + \tfrac{1}{2}\cos{(A-B)}. \tag{50.5}$$

El término a_1 se convierte en

$$\tfrac{1}{2}a_1(\cos 8\omega t + \cos 6\omega t). \tag{50.6}$$

Tenemos entonces dos cosenos, uno con 8 periodos completos en T y el otro con 6. *Ambos promedian cero.* El promedio del término a_1 es por lo tanto cero.

Para el término a_2, encontraríamos $a_2\cos 9\omega t$ y $a_2\cos 5\omega t$, cada uno de los cuales también promedia cero. Para el término a_9, encontraríamos $\cos 16\omega t$ y $\cos{(-2\omega t)}$. Pero $\cos{(-2\omega t)}$ es igual a $\cos 2\omega t$, por lo que los dos tienen promedio cero. Resulta claro que *todos* los términos en a tendrán un promedio cero, *excepto* uno. Y éste es el término a_7. Para él tenemos:

$$\tfrac{1}{2}a_7(\cos 14\omega t + \cos 0). \tag{50.7}$$

El coseno de cero es uno y su promedio, en efecto, es uno. Por lo que resulta que el promedio de todos los términos en a de la ecuación (50.4) es igual a $\frac{1}{2}a_7$.

Los términos en b son aún más fáciles. Cuando multiplicamos por cualquier término en coseno, como $\cos n\omega t$, podemos demostrar siguiendo el mismo método, que *todos* los términos en b tienen el valor promedio de cero.

Vemos que el "truco" de Fourier ha actuado como una criba. Cuando multiplicamos por $\cos 7\omega t$ y promediamos, todos los términos desaparecen excepto a_7 y encontramos que:

$$\text{Promedio de}\,[f(t)\cdot\cos 7\omega t] = a_7/2, \tag{50.8}$$

o

$$a_7 = \frac{2}{T}\int_0^T f(t)\cdot\cos 7\omega t\,dt. \tag{50.9}$$

Dejaremos que el lector demuestre que el coeficiente b_7 se puede obtener multiplicando la ecuación (50.2) por $\text{sen}\,7\omega t$ y promediando ambos miembros. El resultado es:

$$b_7 = \frac{2}{T}\int_0^T f(t)\cdot\text{sen}\,7\omega t\,dt. \tag{50.10}$$

Ahora bien, lo que es verdad para 7 esperamos que sea verdad para cualquier entero. Por ello, podemos resumir nuestra demostración y nuestro resultado en la siguiente forma matemática más elegante. Si m y n son enteros distintos de

cero, y si $\omega = 2\pi/T$, entonces

$$\text{I.}\quad \int_0^T \operatorname{sen} n\omega t \cos m\omega t\, dt = 0. \tag{50.11}$$

$$\left.\begin{aligned} &\text{II.}\quad \int_0^T \cos n\omega t \cos m\omega t\, dt = \\ &\text{III.}\quad \int_0^T \operatorname{sen} n\omega t \operatorname{sen} m\omega t\, dt = \end{aligned}\right\} \begin{cases} 0 & \text{si } n \neq m. \\ T/2 & \text{si } n = m. \end{cases} \tag{50.12}$$

$$\text{IV.}\quad f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega t + \sum_{n=1}^{\infty} b_n \operatorname{sen} n\omega t. \tag{50.13}$$

$$\text{V.}\quad a_0 = \frac{1}{T}\int_0^T f(t)\, dt. \tag{50.14}$$

$$a_n = \frac{2}{T}\int_0^T f(t)\cdot \cos n\omega t\, dt. \tag{50.15}$$

$$b_n = \frac{2}{T}\int_0^T f(t)\cdot \operatorname{sen} n\omega t\, dt. \tag{50.16}$$

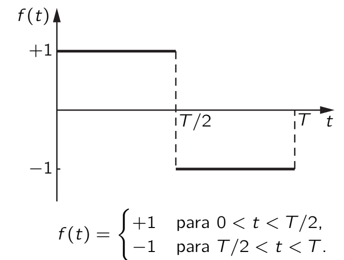
En capítulos anteriores fue conveniente usar notación exponencial para representar un movimiento armónico simple. En lugar de $\cos\omega t$ usamos $\operatorname{Re} e^{i\omega t}$, la parte real de la función exponencial. Hemos usado funciones seno y coseno en este capítulo porque hace un poco más claras quizá las derivaciones. Sin embargo, nuestro resultado final de la ecuación (50.13) se puede escribir en la forma compacta

$$f(t) = \operatorname{Re} \sum_{n=0}^{\infty} \hat{a}_n e^{in\omega t}, \tag{50.17}$$

donde $\hat{a}_n$ es el número complejo $a_n - ib_n$ (con $b_0 = 0$). Si deseamos usar la misma notación en todo el proceso, podemos escribir también

$$\hat{a}_n = \frac{2}{T}\int_0^T f(t) e^{-in\omega t}\, dt \qquad (n \geq 1). \tag{50.18}$$

Ahora sabemos cómo "analizar" una onda periódica en sus componentes armónicas. El procedimiento se llama *análisis de Fourier Fourier* y los términos por separado se llaman componentes de Fourier. *No* hemos demostrado, sin embargo, que una vez que hemos encontrado y sumado todas las componentes de Fourier, obtenemos nuevamente nuestra $f(t)$. Los matemáticos han demostrado, para una amplia gama de funciones y en realidad para todas aquellas que son de interés para los físicos, que si podemos hacer las integrales obtenemos de nuevo $f(t)$. Hay una pequeña excepción. Si la función $f(t)$ es discontinua, esto es, si salta de repente de un valor a otro, la suma de Fourier dará un valor en el punto de ruptura a medio camino entre el valor superior e inferior de la discontinuidad. Así, si tenemos la extraña función $f(t) = 0, 0 \leq t < t_0$ y $f(t) = 1$ para $t_0 \leq t \leq T$, la suma de Fourier dará el valor correcto en todos los puntos *excepto en* t_0, donde tendrá el valor $\frac{1}{2}$ en lugar de 1. No es bastante físico, de todas formas, insistir en que una función debe ser cero *hasta* t_0 y 1 *exactamente en* t_0. Por lo que quizá deberíamos dar la "regla" para los físicos de que cualquier función discontinua (que solamente puede ser una simplificación de una función física *real*) se debería definir con el valor del punto medio de cada discontinuidad. Entonces, cualquier función parecida —con cualquier número finito de tales saltos—, así como todas las otras funciones físicamente interesantes están dadas correctamente por la suma de Fourier.



$$f(t) = \begin{cases} +1 & \text{para } 0 < t < T/2, \\ -1 & \text{para } T/2 < t < T. \end{cases}$$

Figura 50-3. Función onda cuadrada.

Como ejercicio, sugerimos que el lector determine la serie de Fourier de la función mostrada en la figura 50-3. Como la función no se puede escribir en una forma algebraica explícita, no podrán hacer las integrales desde cero hasta T en la forma corriente. Sin embargo, las integrales son fáciles si las separamos en dos

partes, la integral desde cero hasta $T/2$ [donde $f(t) = 1$] y la integral desde $T/2$ hasta T [donde $f(t) = -1$]. El resultado debe ser:

$$f(t) = \frac{4}{\pi}(\operatorname{sen}\omega t + \tfrac{1}{3}\operatorname{sen} 3\omega t + \tfrac{1}{5}\operatorname{sen} 5\omega t + \cdots), \tag{50.19}$$

con $\omega = 2\pi/T$. Encontramos así que nuestra onda cuadrada (con la fase particular escogida) tiene solamente armónicos impares y sus amplitudes están en proporción inversa a sus frecuencias.

Comprobemos que la ecuación (50.19) nos da verdaderamente $f(t)$ de nuevo para algún valor de t. Escojamos $t = T/4$, o sea $\omega t = \pi/2$. Tenemos:

$$f(t) = \frac{4}{\pi}\left(\operatorname{sen}\frac{\pi}{2} + \frac{1}{3}\operatorname{sen}\frac{3\pi}{2} + \frac{1}{5}\operatorname{sen}\frac{5\pi}{2} + \cdots\right) \tag{50.20}$$

$$= \frac{4}{\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \pm \cdots\right). \tag{50.21}$$

La serie* tiene el valor $\pi/4$ y encontramos que $f(t) = 1$.

50-5 Teorema de la energía

La energía de una onda es proporcional al cuadrado de su amplitud. Para una onda de forma compleja, la energía en un periodo será proporcional a $\int_0^T f^2(t)dt$. También podemos relacionar esta energía con los coeficientes de Fourier. Escribimos:

$$\int_0^T f^2(t)\,dt = \int_0^T \left[a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega t + \sum_{n=1}^{\infty} b_n \operatorname{sen} n\omega t\right]^2 dt. \tag{50.22}$$

Cuando desarrollamos el cuadrado del corchete obtendremos todos los dobles productos posibles, como $a_5 \cos 5\omega t \cdot a_7 \cos 7\omega t$ y $a_5 \cos 5\omega t \cdot b_7 \operatorname{sen} 7\omega t$. Hemos demostrado antes, sin embargo [ecuaciones (50.11) y (50.12)], que las integrales de todos esos términos sobre un periodo son cero. Hemos dejado solamente los términos cuadráticos como $a_5^2 \cos^2 5\omega t$. La integral de cualquier coseno o seno al cuadrado sobre un periodo es igual a $T/2$, por lo que obtenemos:

$$\int_0^T f^2(t)\,dt = Ta_0^2 + \frac{T}{2}\,(a_1^2 + a_2^2 + \cdots + b_1^2 + b_2^2 + \cdots)$$

$$= Ta_0^2 + \frac{T}{2}\sum_{n=1}^{\infty}(a_n^2 + b_n^2). \tag{50.23}$$

Esta ecuación se llama "teorema de las enegía" y dice que la energía total de una onda es justamente la suma de las energías de todas las componentes de Fourier. Por ejemplo, aplicando este teorema a la serie (50.19) y puesto que $[f(t)]^2 = 1$ obtenemos:

$$T = \frac{T}{2}\cdot\left(\frac{4}{\pi}\right)^2\left(1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots\right),$$

de donde aprendemos que la suma de los cuadrados de los inversos de los números enteros impares es $\pi^2/8$. De un modo semejante, obteniendo primero la serie de Fourier para la función $f(t) = (t - T/2)^2$ y usando el teorema de la energía podemos probar que $1 + 1/2^4 + 1/3^4 + \cdots$ es $\pi^4/90$, resultado que necesitábamos en el capítulo 45.

---

* La serie se puede evaluar del siguiente modo. Primero observamos que $\int_0^x [dx/(1 + x^2)] = \operatorname{tg}^{-1} x$. En segundo lugar, desarrollamos el integrando en serie: $1/(1+x^2) = 1-x^2+x^4-x^6 \pm \cdots$ Integramos la serie término a término (desde cero a x) para obtener $\operatorname{tg}^{-1} x = x - x^3/3 + x^5/5 - x^7/7 \pm \cdots$ Haciendo $x = 1$, tenemos el resultado establecido, puesto que $\operatorname{tg}^{-1} 1 = \pi/4$.

50-6 Respuestas no lineales

Finalmente, hay un fenómeno importante en la teoría de los armónicos que se debe señalar a causa de su importancia práctica —el de los efectos no lineales—. En todos los sistemas que hemos estado considerando hasta ahora, hemos supuesto que todo era lineal, que las respuestas a las fuerzas, digamos los desplazamientos o las aceleraciones, eran siempre proporcionales a las fuerzas. O bien, que las corrientes en los circuitos eran proporcionales a los voltajes, etc. Queremos considerar ahora casos donde no hay una estricta proporcionalidad. Consideramos, por el momento, algún dispositivo en el que la respuesta, que llamaremos x_{sal} al tiempo t, está determinada por la entrada x_{ent} al tiempo t. Por ejemplo, x_{ent} podría ser la fuerza y x_{sal} podría ser el desplazamiento. O también, x_{ent} podría ser la corriente y x_{sal} el voltaje. Si el dispositivo es lineal, tendríamos:

$$x_{\text{sal}}(t) = Kx_{\text{ent}}(t), \tag{50.24}$$

donde K es una constante independiente de t y de x_{ent}. Supongan, sin embargo, que el dispositivo es casi, pero no exactamente, lineal, de modo que podamos escribir:

$$x_{\text{sal}}(t) = K[x_{\text{ent}}(t) + \epsilon x_{\text{ent}}^2(t)], \tag{50.25}$$

donde ϵ es pequeño frente a la unidad. Las gráficas de la figura 50-4 muestran respuestas lineales y no lineales.

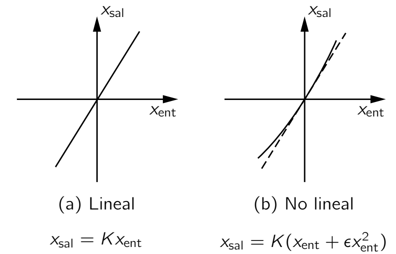


Figura 50-4. Respuestas lineal y no lineal.

Las respuestas no lineales tienen muchas consecuencias prácticas importantes. Ahora discutiremos algunas de ellas. En primer lugar, consideramos lo que sucede si aplicamos una nota pura a la entrada. Hacemos $x_{\text{ent}} = \cos\omega t$. Si representamos x_{sal} en función del tiempo, obtenemos la curva continua que muestra la figura 50-5. La curva discontinua resulta, por comparación, la respuesta de un sistema lineal. Vemos que la salida ya no es una función coseno. Es más aguda arriba y más llana abajo. Decimos que la salida está *distorsionada.* Sabemos, sin embargo, que tal onda ya no es más una nota pura, que tendrá armónicos. Podemos. encontrar cuáles son los armónicos. Usando $x_{\text{ent}} = \cos\omega t$ con la ecuación (50.25), tenemos:

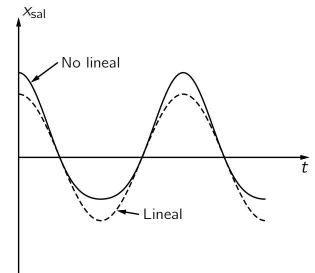


Figura 50-5. Respuesta de un dispositivo no lineal a la entrada $\cos\omega t$. Para comparar se muestra una respuesta lineal.

$$x_{\text{sal}}(t) = K(\cos\omega t + \epsilon\cos^2\omega t). \tag{50.26}$$

De la igualdad $\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$, tenemos:

$$x_{\text{sal}}(t) = K\Big(\cos\omega t + \frac{\epsilon}{2} + \frac{\epsilon}{2}\cos 2\omega t\Big). \tag{50.27}$$

La salida no sólo tiene una componente a la frecuencia fundamental, que estaba presente a la entrada, sino que también tiene algo de su segundo armónico. También ha aparecido a la salida un término constante $K(\epsilon/2)$ que corresponde al corrimiento del valor medio, mostrado en la figura 50-5. El proceso de producir un corrimiento del valor promedio se llama *rectificación.*

Una respuesta no lineal rectificará y producirá armónicos de las frecuencias de su señal de entrada. Aunque la falta de linealidad que supusimos produjera solamente segundos armónicos, la falta de linealidad de orden más alto —aquellas que tienen términos como x_{ent}^3 y x_{ent}^4 por ejemplo— producirán armónicos más altos que el segundo.

Otro efecto que resulta de una respuesta no lineal es la *modulación.* Si nuestra función de entrada contiene dos (o más) notas puras, la salida tendrá no solamente sus armónicos, sino también otras componentes de frecuencia. Hagamos $x_{\text{ent}} = A\cos\omega_1 t + B\cos\omega_2 t$, donde ahora ω_1 y ω_2 *no* se pretende que estén en relación armónica. Además del término lineal (que es K por la señal) tendremos una componente a la salida dada por:

$$x_{\text{sal}} = K\epsilon(A\cos\omega_1 t + B\cos\omega_2 t)^2 \tag{50.28}$$

$$= K\epsilon(A^2\cos^2\omega_1 t + B^2\cos^2\omega_2 t + 2AB\cos\omega_1 t\cos\omega_2 t). \tag{50.29}$$

Los dos primeros términos dentro del paréntesis en la ecuación (50.29) son justamente los que dieron los términos constantes y los segundos armónicos que encontramos antes. El último término es nuevo.

Podemos considerar este nuevo "término cruzado" $AB\cos\omega_1 t\cos\omega_2 t$ de dos modos. En primer lugar, si las dos frecuencias son muy diferentes (por ejemplo, si ω_1 es mucho mayor que ω_2) podemos considerar que el término cruzado representa una oscilación cosenoidal de amplitud variable. Esto es, podemos considerar los factores en esta forma:

$$AB\cos\omega_1 t\cos\omega_2 t = C(t)\cos\omega_1 t, \tag{50.30}$$

con

$$C(t) = AB\cos\omega_2 t. \tag{50.31}$$

Decimos que la amplitud de $\cos\omega_1 t$ está modulada con la frecuencia ω_2.

Alternativamente, podemos escribir el término cruzado en otra forma:

$$AB\cos\omega_1 t\cos\omega_2 t = \frac{AB}{2}\,[\cos{(\omega_1+\omega_2)t} + \cos{(\omega_1-\omega_2)t}]. \tag{50.32}$$

Ahora diríamos que se han producido dos *nuevas* componentes, una a la frecuencia *suma* $(\omega_1+\omega_2)$, otra a la frecuencia *diferencia* $(\omega_1-\omega_2)$.

Tenemos dos modos diferentes, pero equivalentes, de considerar el mismo resultado. En el caso especial en que $\omega_1 \ll \omega_2$, podemos relacionar estos dos puntos de vista diferentes señalando que como $(\omega_1+\omega_2)$ y $(\omega_1-\omega_2)$ son próximas esperaríamos observar pulsaciones entre ellas. Pero estas pulsaciones tienen solamente el efecto de *modular* la amplitud de la frecuencia *promedio* ω_1 en la mitad de la frecuencia diferencia $2\omega_2$. Vemos entonces por qué las dos descripciones son equivalentes.

En resumen, hemos encontrado que una respuesta no lineal produce varios efectos: rectificación, generación de armónicos y modulación o generación de componentes con las frecuencias suma y diferencia.

Deberíamos notar que todos estos efectos [ecuación (50.29)] son proporcionales no solamente al coeficiente de falta de linealidad ϵ, sino también al producto de dos amplitudes —bien A^2, B^2 o AB—. Es de esperar que estos efectos sean mucho más importantes para las señales *fuertes* que para las débiles.

Los efectos que hemos estado describiendo tienen muchas aplicaciones prácticas. En primer lugar, respecto al sonido, se considera que el oído no es lineal. Se considero que esto explica el hecho de que con los sonidos fuertes tenemos la sensación de que *oímos* armónicos y también frecuencias suma y diferencia aunque las ondas sonoras contengan solamente notas puras.

Los componentes que se usan en equipos de reproducción de sonido —amplificadores, altoparlantes, etc.— siempre tienen algo de falta de linealidad. Producen distorsiones en el sonido —generan armónicos, etc.— que no estaban presentes en el sonido original. El oído oye estas nuevas componentes, que son aparentemente objetables. Es por esta razón que los equipos "Hi-Fi" están diseñados para que sean lo más lineales posible. (Por qué la falta de linealidad del *oído no* es objetable del mismo modo, o cómo sabemos que la falta de linealidad está en el *altoparlante* y no en el *oído*, ¡no está claro!)

Las faltas de linealidad son muy *necesarias* y, en la realidad, se hacen intencionalmente grandes en ciertas partes de los equipos radiotransmisores o receptores. En un transmisor AM la señal "voz" (con frecuencia de algunos kilociclos por segundo) se combina con la señal "portadora" (con una frecuencia de algunos megaciclos por segundo) en un circuito no lineal llamado *modulador*, para producir la oscilación modulada que se transmite. En el receptor, las componentes de la señal recibida alimentan un circuito no lineal que combina las frecuencias suma y diferencia de la portadora modulada para generar de nuevo la señal voz.

Cuando discutimos la transmisión de la luz, supusimos que las oscilaciones inducidas de cargas eran proporcionales al campo eléctrico de la luz —que la respuesta era lineal—. Esto es en verdad una aproximación muy buena. Ha sido solamente en los últimos años que se han diseñado fuentes de luz (láseres)

que producen una intensidad de luz lo bastante fuerte como para que se puedan observar los efectos no lineales. Ahora es posible generar armónicos de las frecuencias de la luz. Cuando una luz roja intensa atraviesa un trozo de vidrio, ¡sale un poquito de luz azul —segundo armónico—!

51

Ondas

51-1 Ondas de proa

Aunque hemos terminado nuestro análisis cuantitativo de las ondas, este capítulo adicional sobre el tema está destinado a dar cierta visión cualitativa de diversos fenómenos asociados con las ondas, los cuales son demasiado complicados como para analizarlos en detalle aquí. Como hemos estado tratando ondas durante varios capítulos, un nombre más apropiado del tema sería "algunos de los fenómenos más complejos asociados con las ondas".

El primer tópico a discutir se refiere a los efectos producidos por una fuente de ondas que se mueve a una velocidad mayor que la de la onda, o velocidad de fase. Consideremos primero ondas que tienen una velocidad definida, como el sonido y la luz. Si tenemos una fuente de sonido moviéndose a una velocidad mayor que la del sonido, ocurre algo así: supongan que en cierto momento se genera una onda sonora en la fuente en el punto x_1 de la figura 51-1; entonces, en el instante siguiente, a medida que la fuente se traslada hasta x_2 la onda proveniente de x_1 se expande hasta un radio r_1 menor que la distancia recorrida por la fuente, y por supuesto, otra onda sale de x_2. Cuando la fuente de sonido se ha movido aún más, hasta x_3, y una onda está saliendo allí, la onda proveniente de x_2 se ha expandido ahora hasta r_2 y la proveniente de x_1 hasta r_3. Naturalmente, la cosa ocurre continuamente, no a saltos, y en consecuencia tenemos una serie de ondas circulares con una recta tangente común que pasa por el centro de la fuente. Vemos que en vez de una fuente que genera ondas esféricas, como ocurriría si estuviese quieta, genera un frente de onda que forma un cono en tres dimensiones, o un par de rectas en dos dimensiones. El ángulo del cono es muy fácil de hallar. Durante un intervalo dado de tiempo la fuente se mueve cierta distancia, digamos que $x_3 - x_1$ proporcional a v, la velocidad de la fuente. Mientras tanto, el frente de onda se ha movido una distancia r_3 proporcional a c_w, la velocidad de la onda. En consecuencia, es evidente que el semiángulo de abertura tiene un seno igual al cociente entre la velocidad de las ondas y la velocidad de la fuente, y este seno tiene solución solamente si c_w es menor que v, o sea que la velocidad del objeto es mayor que la de la onda:

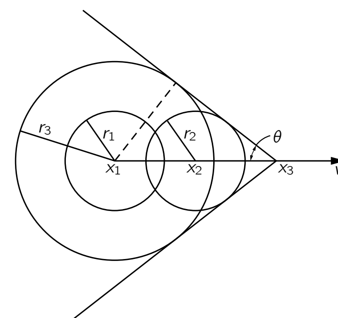


Figura 51-1. El frente de onda de choque yace sobre un cono con vértice en la fuente y semiapertura $\theta = \operatorname{sen}^{-1} c_w/v$.

$$\operatorname{sen}\theta = \frac{c_w}{v}. \tag{51.1}$$

Entre paréntesis, aunque hemos supuesto implícitamente que es necesario tener una *fuente* de sonido, resulta, y esto es muy interesante, que una vez que el objeto se mueve más rápidamente que el sonido, *producirá* sonido. Esto es, no es necesario que tenga un carácter vibracional con cierto tono. Cualquier objeto que se mueva en un medio con una velocidad mayor que la velocidad a la cual el medio transporta ondas, generará ondas a cada lado, automáticamente, por el simple movimiento. Esto es sencillo en el caso del sonido, pero también ocurre en el caso de la luz. A primera vista se podría pensar que nada se puede mover más rápido que la luz. Sin embargo, la luz tiene en el vidrio una velocidad de fase menor que la velocidad de la luz en el vacío y es posible disparar una partícula cargada de energía muy alta a través de un bloque de vidrio de modo que la velocidad de la partícula sea cercana a la velocidad de la luz en el vacío, mientras que la velocidad de la luz en el vidrio puede ser sólo $\frac{2}{3}$ de la velocidad de la luz en el vacío. Una partícula que se mueve más rápido que la luz en el medio producirá una onda cónica de luz con vértice en la fuente, como la onda de la estela de un barco (que en realidad proviene del mismo efecto). Midiendo

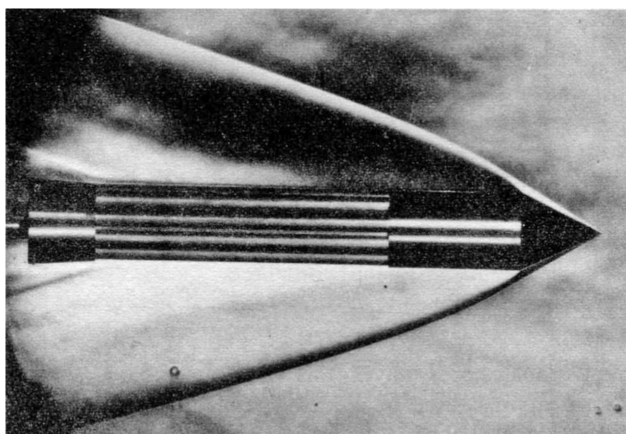


Figura 51-2. Onda de choque inducida en un gas por un proyectil que se mueve más rápido que el sonido.

el ángulo del cono podemos determinar la velocidad de la partícula y es uno de los métodos para determinar su energía en la investigación de alta energía. Todo lo que se necesita medir es la dirección de la luz.

Esta luz se llama a veces *radiación de de Cherenkov* porque Cherenkov fue el primero en observarla. Franck y Tarnm analizaron teóricamente cuál debería ser la intensidad de esta luz. El Premio Nobel de Física 1958 fue otorgado conjuntamente a los tres por este trabajo.

La situación correspondiente en el caso del sonido está ilustrada en la figura 51-2, que es una fotografía de un objeto moviéndose a través de un gas con una velocidad mayor que la velocidad del sonido. Las variaciones de presión producen una variación del índice de refracción y con un sistema óptico apropiado es posible hacer visibles los bordes de las ondas. Vemos que el objeto que se mueve más rápido que el sonido produce, verdaderamente, una onda cónica. Pero una observación más atenta revela que en realidad la superficie es curva. Es recta asintóticamente, pero es curva cerca del vértice, y tenemos ahora que discutir cómo puede ser esto, lo cual nos lleva al segundo tópico de este capítulo.

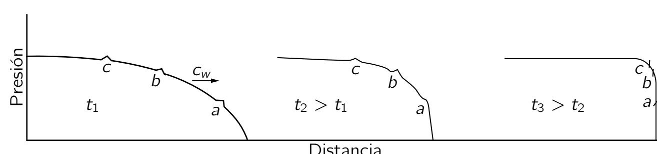


Figura 51-3. "Instantáneas" del frente de onda en instantes sucesivos de tiempo.

51-2 Ondas de choque

A menudo, la velocidad de las ondas depende de la amplitud, y en el caso del sonido la velocidad depende de la amplitud en la siguiente forma. Un objeto que se mueve en el aire, tiene que apartar a éste de su camino, por lo que la perturbación producida en este caso es una especie de escalón de presión, siendo la presión más alta detrás del frente de onda que en la región no perturbada que aún no ha alcanzado la onda (que se mueve a velocidad normal, digamos). Pero el aire dejado atrás después que pasa el frente de onda se ha comprimido adiabáticamente y en consecuencia la temperatura aumenta. Ahora bien, la velocidad del sonido aumenta con la temperatura, por lo que en la región que está detrás del salto, la velocidad es mayor que en el aire que está al frente. Esto implica que cualquier otra perturbación que se haga detrás de este escalón, digamos que empujando continuamente el cuerpo, o cualquier otra perturbación, se moverá más rápido que el frente, aumentando la velocidad con el incremento de presión. La figura 51-3 ilustra la situación, con algunas pequeñas protuberancias de presión agregadas al perfil de presión para ayudar a visualizar. Vemos que

las regiones de mayor presión en la parte posterior sobrepasan el frente con el transcurrir del tiempo, hasta que finalmente la onda de compresión desarrolla un frente bien definido. Si la intensidad es muy alta, "finalmente" significa repentinamente. Si es más bien débil, tarda mucho tiempo, tanto que puede ocurrir que el sonido se difunda y se extinga antes de que tenga tiempo de desarrollar este frente.

Los sonidos que hacemos al hablar son extremadamente débiles respecto a la presión atmosférica —sólo uno en un millón más o menos—. Pero para variaciones de presión del orden de una atmósfera, la velocidad de la onda aumenta en alrededor de 20 % y el frente de onda se hace bien definido con una rapidez correspondientemente alta. Es de presumir que nada ocurre en la naturaleza *infinitamente* rápido, y lo que llamamos frente "bien definido" tiene, en realidad, un pequeñísimo espesor, no es infinitamente abrupto. Las distancias sobre las cuales varía son del orden de un camino libre medio, para las cuales la teoría de la ecuación de onda comienza a fallar porque no hemos considerado la estructura del gas.

Ahora bien, refiriéndonos de nuevo a la figura 51-2, vemos que se puede comprender la curvatura si nos damos cuenta de que cerca del vértice las presiones son mayores de lo que son más atrás, por lo que el ángulo θ es mayor. Esto es, la curva es el resultado del hecho de que la velocidad depende de la intensidad de la onda. Por consiguiente, la onda producida por la explosión de una bomba atómica viaja por un corto tiempo a una velocidad mucho más alta que la del sonido, hasta que está tan alejada que es debilitada por la difusión, a tal punto que el pico de presión es pequeño comparado con la presión atmosférica. La velocidad de la perturbación se aproxima a la velocidad del sonido en el gas donde se propaga. (Entre paréntesis, siempre resulta que la velocidad de la onda de choque es más alta que la velocidad del sonido en el gas que está delante, pero es más baja que la velocidad del sonido en el gas que está atrás. Es decir, los impulsos provenientes de atrás llegarán al frente, pero éste avanza hacia el medio en el cual se mueve con velocidad mayor que la velocidad normal de las señales. Así, pues, no se puede saber acústicamente que la onda está viniendo hasta que es demasiado tarde. La luz proveniente de la bomba llega primero, pero no se puede saber que viene la onda de choque hasta que llega, porque no hay señal sonora que la preceda.)

Este fenómeno, este apilamiento de ondas, es muy interesante, y el punto fundamental del cual depende es que, después que hay una onda, la velocidad de la onda resultante debe ser mayor. El siguiente es otro ejemplo del mismo fenómeno. Consideren agua corriendo por un largo canal de ancho y profundidad finitos. Si a lo largo del canal se mueve un pistón, o una pared a lo ancho del canal, suficientemente rápido, el agua se amontona, como la tierra frente a una topadora niveladora. Supongan ahora que la situación es como muestra la figura 51-4, con un escalón repentino en la altura del agua en algún punto del canal. Se puede demostrar que las ondas largas en un canal viajan más rápido en aguas profundas que en aguas superficiales. En consecuencia, cualquier nueva perturbación o irregularidad en la energía suministrada por el pistón se adelanta y se amontona al frente. De nuevo, lo que tenemos al final es simplemente agua con un frente bien delineado, teóricamente. Sin embargo, como muestra la figura 51-4, hay complicaciones. Se ha fotografiado una onda que recorre un canal, el pistón está en el extremo derecho del canal. Al principio puede que haya aparecido como una onda que se porta bien, como sería de esperar, pero más adelante en el canal se ha vuelto más y más puntiaguda hasta que ocurrió lo que se ha fotografiado. La superficie se revuelve terriblemente a medida que caen porciones de agua, pero esencialmente es una elevación bastante aguda sin ninguna perturbación del agua adelante.

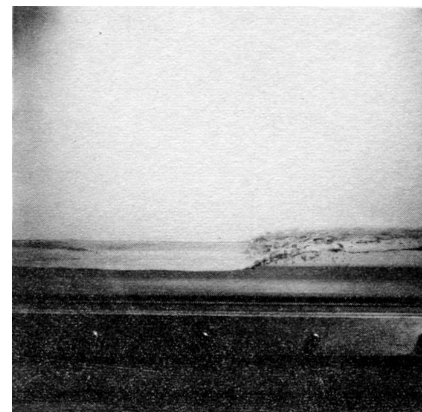


Figura 51-4.

En realidad, el agua es mucho más complicada que el sonido. Sin embargo, sólo para dar un ejemplo de esto, trataremos de analizar la velocidad de esta *gran ola en un canal*. Este punto no tiene ninguna importancia básica para nuestros fines —no es una gran generalización—, es solamente para dar un ejemplo de que las leyes de la mecánica que ya conocemos son capaces de explicar el fenómeno.

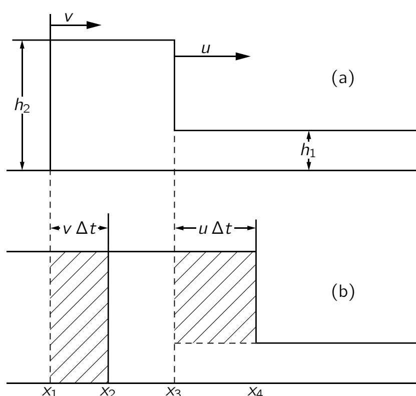


Figura 51-5. Dos secciones transversales de una gran ola en un canal, siendo (b) un intervalo Δt posterior a (a).

Imaginen por un momento que el agua tiene más o menos el aspecto de la figura 51-5(a), que el agua a la altura más alta h_2 se mueve con velocidad v y que el frente se mueve con velocidad u hacia el agua no perturbada que está a la altura h_1. Querríamos determinar la velocidad a la que se mueve el frente. En un tiempo Δt, un plano vertical inicialmente en x_1 se mueve a una distancia $v\,\Delta t$ hasta x_2 mientras el frente de la onda se ha movido $u\,\Delta t$.

Apliquemos ahora las ecuaciones de conservación de la materia y del momentum. En primer lugar la materia: vemos que por unidad de ancho del canal la cantidad $h_2 v\,\Delta t$ de materia que ha pasado por x_1 (que se muestra rayada) es compensada por la otra región rayada de magnitud $(h_2 - h_1)u\,\Delta t$. Luego, dividiendo entre Δt, $vh_2 = u(h_2-h_1)$. Esto no nos basta, porque aunque tenemos h_2 y h_1, no conocemos ni u ni v, estamos tratando de obtener las dos.

Ahora, el paso siguiente es emplear la conservación del momentum. No hemos discutido los problemas de la presión del agua, ni nada de hidrodinámica, pero de todos modos está claro que la presión del agua a una profundidad dada es precisamente la suficiente como para sostener la columna de agua que está por encima. En consecuencia, la presión del agua es igual a ρ, densidad del agua, multiplicada por g, multiplicada por la profundidad desde la superficie. Como la presión aumenta linealmente con la profundidad, la presión media sobre el plano que está en x_1, digamos, es $\frac{1}{2}\rho g h_2$, que es también la fuerza media por unidad de ancho y por unidad de altura, que empuja el plano hacia x_2. Por lo tanto, multiplicamos por otra h_2 para obtener la fuerza total que se ejerce sobre el agua que empuja desde la izquierda. Además, está también la presión en el agua de la derecha que ejerce una fuerza opuesta sobre la región en cuestión, que por el mismo tipo de análisis es $\frac{1}{2}\rho g h_1^2$. Ahora debemos equilibrar las fuerzas con la rapidez de variación del momentum. Así, pues, tenemos que calcular cuánto momentum en exceso hay en la situación (b) de la figura 51-5 respecto al que había en (a). Vemos que la masa adicional que ha adquirido la velocidad v es simplemente $\rho h_2 u\,\Delta t - \rho h_2 v\,\Delta t$ (por unidad de ancho). Multiplicando esta masa por v se obtiene el momentum adicional a igualar con el impulso $F\,\Delta t$:

$$(\rho h_2 u\,\Delta t - \rho h_2 v\,\Delta t)v = (\tfrac{1}{2}\rho g h_2^2 - \tfrac{1}{2}\rho g h_1^2)\,\Delta t.$$

Si eliminamos v de esta ecuación sustituyendo $vh_2 = u(h_2 - h_1)$ que ya hemos encontrado, y simplificamos, obtenemos finalmente que $u^2 = gh_2(h_1 + h_2)/2h_1$.

Si la diferencia de alturas es muy pequeña, de modo que h_1 y h_2 son casi iguales, esto nos dice que la velocidad $= \sqrt{gh}$. Como veremos más adelante, esto es verdad únicamente si la longitud de onda de la onda es mayor que la profundidad del canal.

También podríamos haber hecho lo mismo para las ondas sonoras —incluyendo la conservación de la energía interna en lugar de conservación de la entropía, porque la onda de choque es irreversible—. Verdaderamente, al fijarse si se conserva la energía en el problema de la gran ola, se encuentra que la energía no se conserva. Si la diferencia de alturas es pequeña, se conserva casi perfectamente, pero tan pronto como la diferencia de alturas se hace muy apreciable, hay una pérdida neta de energía. Esto se manifiesta como caída de agua con el consiguiente revoltijo mostrado en la figura 51-4.

En las ondas de choque hay una correspondiente pérdida aparente de energía, desde el punto de vista de las reacciones adiabáticas. La energía de la onda sonora, detrás de la onda de choque, se transforma en calentamiento del gas después que pasa la onda de choque, lo cual corresponde al revoltijo del agua en la ola. Al calcular esto resulta para el caso del sonido que son necesarias tres ecuaciones para la solución, y la temperatura detrás de la onda de choque no es igual a la temperatura al frente, como hemos visto.

Si tratamos de hacer una ola invertida ($h_2 < h_1$) encontramos que la pérdida de energía por segundo es negativa. Como no hay energía disponible en ninguna parte, esa ola no se puede mantener, es inestable. Si formáramos una onda de ese tipo se aplanaría porque la dependencia de la velocidad con la altura que daba lugar al frente afilado en el caso discutido, tendría ahora el efecto opuesto.

51-3 Ondas en sólidos

La siguiente clase de ondas a estudiar es la de las ondas más complicadas en sólidos. Ya hemos tratado ondas en gases y en líquidos y hay una analogía directa con una onda sonora en un sólido. Si se aplica a un sólido un golpe repentino, se comprime. Resiste la compresión y se establece una onda análoga al sonido. Sin embargo, hay otro tipo de onda posible en un sólido, y que no es posible en un fluido. Si se deforma un sólido tirando de él lateralmente (llamada *deformación de corte*) trata de regresar a su forma original. Esto es por definición lo que distingue un sólido de un líquido: si deformamos un líquido (internamente), lo mantenemos un minuto para que se calme y luego lo soltamos, se queda así, pero si tomamos un sólido y lo deformamos, como deformar lateralmente un pedazo de gelatina, y luego lo soltamos, vuelve a como estaba, iniciando una onda *de corte* que viaja de la misma manera que las compresiones. En todos los casos la velocidad de la onda de corte es menor que la de las ondas longitudinales. Las ondas de cortes son parecidas, en lo que respecta a sus polarizaciones, a las ondas luminosas. El sonido no tiene polarización, es simplemente como una onda de presión. La luz tiene una orientación característica perpendicular a su dirección de propagación.

En un sólido, las ondas son de ambos tipos. Primero, hay una onda de compresión análoga al sonido, que corre a una velocidad. Si el sólido no es cristalino se propagará una onda de corte polarizada en cualquier dirección a una velocidad característica. (Por supuesto que todos los sólidos son cristalinos, pero si usamos un bloque hecho de microcristales en todas las orientaciones posibles, las anisotropías cristalinas se promedian y desaparecen.)

Otra pregunta interesante referente a las ondas sonoras es la siguiente: ¿qué ocurre si la longitud de onda en el sólido se acorta más, y más, y más? ¿Hasta dónde se puede acortar? Es interesante que no se puede hacer más corta que el espaciamiento entre los átomos, porque si se supone que hay una onda en la que un punto va para arriba y el siguiente para abajo, etc., la longitud de onda más corta posible es evidentemente el espaciamiento atómico. En términos de modos de oscilación decimos que hay modos longitudinales y transversales, modos de ondas largas y de ondas cortas. Cuando consideramos longitudes de onda comparables al espaciamiento entre átomos, las velocidades ya no son constantes, hay un efecto dispersivo en el que la velocidad no es independiente del número de onda. Pero en última instancia, el modo de vibración más alto de ondas transversales sería aquel en que cada átomo está haciendo lo contrario de los átomos vecinos.

Ahora bien, desde el punto de vista de los átomos, la situación es como los dos péndulos de que estuvimos hablando, para los cuales hay dos modos: uno en el que ambos van juntos y otro en el que van en sentidos opuestos. Es posible analizar las ondas en sólidos de otra manera, en términos de un sistema de osciladores armónicos acoplados, como una cantidad enorme de péndulos, siendo el modo más alto tal que oscilan en forma opuesta y habiendo modos más bajos con diferentes relaciones de sincronización.

Las longitudes de onda más cortas lo son tanto que generalmente no son accesibles técnicamente. Sin embargo, son de gran interés, porque en la teoría termodinámica de un sólido, las propiedades térmicas de un sólido, por ejemplo los calores específicos, se pueden analizar en función de las propiedades de las ondas sonoras cortas. Yendo al extremo de ondas sonoras de longitud de onda siempre más corta, se llega necesariamente a los movimientos individuales de los átomos. En última instancia, las dos cosas son lo mismo.

Un ejemplo muy interesante de ondas sonoras en un sólido, tanto longitudinales como transversales, es el de las ondas que hay en la Tierra sólida. No sabemos quién hace los ruidos, pero de tiempo en tiempo, dentro de la Tierra hay terremotos —alguna roca se desliza sobre alguna otra—. Esto es como un pequeño ruido. Así, de esa fuente parten ondas como las sonoras de longitud de onda mucho más larga de lo que comúnmente se considera en ondas sonoras, pero aún son ondas sonoras y se propagan por la Tierra. Sin embargo, la

Tierra no es homogénea y las propiedades de presión, densidad, compresibilidad, etc., varían con la profundidad y en consecuencia la velocidad varía con la profundidad. Entonces las ondas no viajan en línea recta —hay una especie de índice de refracción y siguen líneas curvas—. Las ondas longitudinales y las transversales tienen velocidad distinta, por lo que hay soluciones diferentes para las distintas velocidades. En consecuencia, si colocamos un sismógrafo en algún sitio y observamos cómo se sacude después de que ha habido un terremoto en alguna parte, no obtenemos simplemente una agitación irregular. Podríamos obtener una agitación, luego un aquietamiento, y después otra agitación —lo que ocurre depende de la ubicación—. Si estuviéramos lo bastante cerca, primero recibiríamos ondas longitudinales provenientes de la perturbación, y luego, unos instantes más tarde, ondas transversales ya que viajan más lentamente. Midiendo la diferencia de tiempo entre las dos, podemos decir a qué distancia está el terremoto si conocemos lo suficiente acerca de las velocidades y de la composición de las regiones internas afectadas.

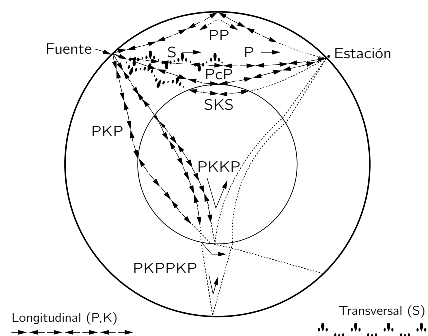


Figura 51-6. Representación esquemática de la Tierra, que muestra el recorrido de ondas sonoras longitudinales y transversales.

La figura 51-6 muestra un ejemplo del diagrama de comportamiento de ondas terrestres. Las dos clases de ondas están representadas por símbolos distintos. Si hubiera un terremoto en el lugar marcado “fuente”, las ondas transversales y las longitudes llegarían en tiempos diferentes a la estación por las rutas más directas, y también habría reflexiones en las discontinuidades, dando lugar a otros recorridos y tiempos. Resulta que la Tierra tiene un núcleo que no transporta ondas transversales. Si la estación está opuesta a la fuente, las ondas transversales siguen llegando, pero el tiempo no es el que corresponde. Lo que ocurre es que la onda transversal llega al núcleo y siempre que las ondas transversales llegan a una superficie oblicua entre dos materiales se generan dos nuevas ondas, una transversal y una longitudinal. Pero una onda transversal no se propaga dentro del núcleo terrestre (o por lo menos no hay evidencia de ello, sólo la hay de ondas longitudinales), sale de nuevo en ambas formas y llega a la estación.

Es a partir del comportamiento de estas ondas de terremotos que se ha determinado que las ondas transversales no se pueden propagar dentro del círculo interior. Esto significa que el centro de la Tierra es líquido en el sentido de que en él no se pueden propagar ondas transversales. La única manera de saber lo que hay dentro de la Tierra es estudiando los terremotos. Así, pues, usando una gran cantidad de observaciones de muchos terremotos en diferentes estaciones ha sido posible obtener los detalles —se conoce la velocidad, las curvas, etc.—. Sabemos cuáles son las velocidades de diversos tipos de ondas a cada profundidad. Conociendo esto, en consecuencia, es posible determinar los modos normales de la Tierra, porque conocemos la velocidad de propagación de las ondas sonoras —en otras palabras, las propiedades elásticas de ambos tipos de ondas a cada profundidad—. Supongan que se deformase la Tierra en forma de elipsoide y se la soltara. Es simplemente cuestión de superponer ondas que viajan por el elipsoide para determinar el periodo y las formas de un modo de oscilación libre. Hemos calculado que si hay una perturbación, hay un montón de modos, desde el más bajo, que es elipsoidal, hasta los modos más altos con más complejidad.

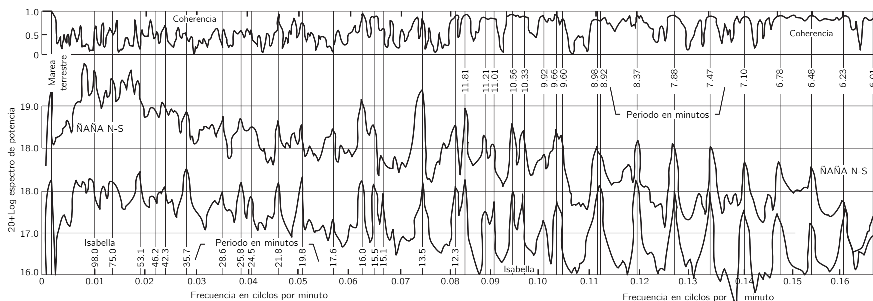


Figura 51-7. Potencia en función de la frecuencia registrada en sismógrafos en Ñaña, Perú, e Isabella, California. La coherencia es una medida del acoplamiento entre las estaciones (Tomada de Benioff, Press y Smith, *J. Geoph. Research* **66**, 605 (1961)).

El terremoto chileno de mayo de 1960 hizo un “ruido” lo suficientemente fuerte como para que las señales dieran vuelta a la Tierra varias veces, y hubo nuevos sismógrafos de gran delicadeza terminados justo a tiempo para determinar la frecuencia de los modos fundamentales de la Tierra y compararlos con los valores calculados a partir de la teoría del sonido con las velocidades conocidas, medidas en terremotos independientes. El resultado de este experimento está ilustrado en la figura 51-7, que es una representación de la intensidad de la señal en función de la frecuencia de su oscilación (un *análisis de Fourier*). Observen que a ciertas frecuencias determinadas se recibe mucho más que a otras frecuencias; hay máximos muy definidos. Éstas son las frecuencias naturales de la Tierra, porque son las frecuencias principales a las cuales puede oscilar la Tierra. En otras palabras, si el movimiento entero de la Tierra está construido con muchos modos diferentes, sería de esperar que en cada estación se obtuvieran sacudidas irregulares que indican superposición de muchas frecuencias. Si analizamos esto en términos de frecuencias, tendríamos que poder encontrar las frecuencias características de la Tierra. Las líneas oscuras verticales de la figura son las frecuencias calculadas, encontramos un acuerdo notable, acuerdo debido a que la teoría del sonido es correcta para el interior de la Tierra.

Hay un punto muy curioso que revela la figura 51-8, la cual muestra una medida muy cuidadosa, con mejor resolución del modo de vibración más bajo: el modo elipsoidal de la Tierra. Observen que no hay un máximo único, sino doble: 54.7 y 53.1 minutos —ligeramente diferentes—. La causa de las dos frecuencias diferentes no se conocía en la época en que se midió, aunque puede que haya sido encontrada desde entonces. Hay por lo menos dos explicaciones posibles. Una sería que puede haber asimetría en la distribución de la Tierra, lo cual daría lugar a dos modos similares. La otra posibilidad, aún más interesante, es ésta: imaginen las ondas dando vuelta a la Tierra en dos direcciones partiendo de la fuente. Las velocidades no serán iguales debido a los efectos de la rotación de la Tierra en las ecuaciones de movimiento, los cuales no se han tenido en cuenta al hacer el análisis. En un sistema rotante, el movimiento está modificado por las fuerzas de coriolis, y éstas podrían originar el desdoblamiento observado.

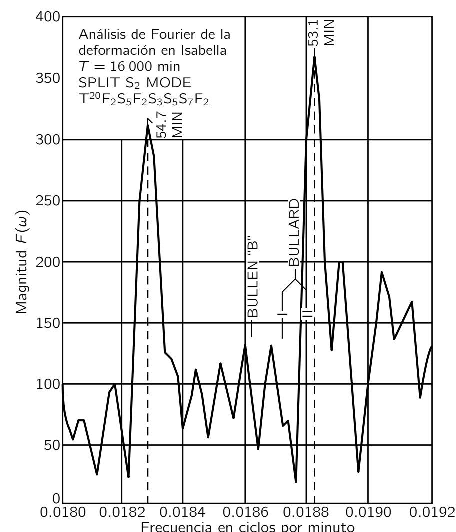


Figura 51-8. Análisis con alta resolución de uno de los registros de los sismógrafos, mostrando un doblete espectral.

Respecto al método con el cual se han analizado estos terremotos, lo que se obtiene en el sismógrafo no es una curva de amplitud en función de la frecuencia, sino desplazamiento en función del tiempo, que siempre es una traza muy irregular. Para hallar la cantidad de todas las ondas sinusoidales diferentes para todas las frecuencias distintas, sabemos que el truco es multiplicar los datos por una onda sinusoidal de frecuencia determinada a integrar, es decir, promediar, y en el promedio desaparecen todas las otras frecuencias. Las figuras eran, pues, representaciones de las integrales encontradas multiplicando los datos por ondas sinusoidales de diferentes ciclos por minuto e integrados.

51-4 Ondas de superficie

Ahora, las siguientes ondas de interés, que cualquiera puede ver fácilmente y que se usan comúnmente como ejemplo de ondas en los cursos elementales, son las ondas de agua. Como veremos inmediatamente, son el peor ejemplo posible porque no son de ninguna manera como el sonido y la luz, tienen todas las complicaciones que las ondas pueden tener. Comencemos con ondas de agua largas en aguas profundas. Si se considera el océano como infinitamente profundo y se hace una perturbación en la superficie, se generan ondas. Ocurre toda clase de movimientos irregulares, pero el movimiento tipo sinusoidal con una perturbación muy pequeña, podría parecerse a las ondas lisas comunes del océano que vienen hacia la costa. Ahora bien, en esa onda el agua, naturalmente, permanece quieta en promedio, pero la onda se mueve. ¿Cómo es el movimiento?, ¿es transversal o longitudinal? Ni lo uno ni lo otro, no es ni transversal ni longitudinal. Aunque en un lugar dado el agua sea alternadamente un valle y una cresta, no puede estar simplemente moviéndose hacia arriba o hacia abajo debido a la conservación del agua. Esto es, si va para abajo, ¿a dónde va a ir el agua? El agua es esencialmente incompresible. La velocidad de compresión de las ondas —es decir, el sonido en

el agua— es mucho, pero mucho más alta, y no estamos considerando eso ahora. Como el agua es incompresible en esta escala, cuando una loma baja, el agua tiene que salir del lugar. Lo que ocurre es que las partículas de agua cercanas a la superficie se mueven aproximadamente en circunferencias. Cuando se acerca una ola lisa, una persona flotando en un neumático puede mirar un objeto cercano y ver que está moviéndose en una circunferencia. Así, es una mezcla de longitudinal y transversal, para aumentar la confusión. A profundidades mayores en el agua, los movimientos son circunferencias más pequeñas hasta que, razonablemente abajo, no queda nada del movimiento (figura 51-9).

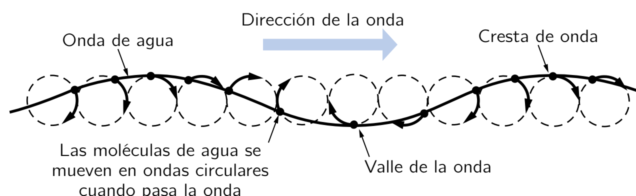


Figura 51-9. Las ondas en aguas profundas están formadas de partículas que se mueven en circunferencias. Notar los desfasajes sistemáticos de circunferencia en circunferencia. ¿Cómo se movería un objeto flotante?

Hallar la velocidad de esas ondas es un problema interesante: debe ser alguna combinación de la densidad del agua, la aceleración de la gravedad, que es la fuerza de restauración que genera las ondas, y posiblemente la longitud de onda y la profundidad. Si tomamos el caso en que la profundidad se hace infinita, ya no dependerá de la profundidad. Para cualquier fórmula que obtengamos, la velocidad de las fases de las ondas debe combinar los diversos factores para dar las dimensiones apropiadas, y si intentamos esto de varias maneras sólo encontramos una manera de combinar la densidad, g y λ, para formar una velocidad: $\sqrt{g\lambda}$, que de ningún modo incluye la densidad. En realidad, esta fórmula para la velocidad de fase no es del todo correcta. Un análisis dinámico completo, en el cual no entraremos, muestra que los factores son los que tenemos, exceptuando $\sqrt{2\pi}$:

$$v_{\text{fase}} = \sqrt{g\lambda/2\pi} \text{ (para ondas gravitacionales).}$$

Es interesante que las ondas largas van más rápido que las cortas. Por lo tanto, si una lancha crea ondas a lo lejos, porque hay un piloto de carros deportivos en una lancha a motor que está pasando, entonces, un poco después, las ondas llegarán a la costa con chapaleos lentos al principio y luego más y más rápidos, porque las primeras ondas que llegan son largas. Las ondas se acortan más y más a medida que pasa el tiempo porque las velocidades son proporcionales a la raíz cuadrada de la longitud de onda.

Se podría objetar: "Eso no está bien, ¡tenemos que considerar la velocidad *de grupo* en el cálculo". Claro que sí. La fórmula para la velocidad de fase no nos dice qué es lo que va a llegar primero, la velocidad de grupo es la que nos lo dice. Así, pues, tenemos que calcular la velocidad de grupo, y se deja como problema demostrar que es la mitad de la velocidad de fase, suponiendo que la velocidad es proporcional a la raíz cuadrada de la longitud de onda, que es todo lo que se necesita. Además, la velocidad de grupo también es proporcional a la raíz cuadrada de la longitud de onda. ¿Cómo es posible que la velocidad de grupo sea la mitad de la de fase? Si se examina el manojo de ondas que hace una lancha en movimiento, siguiendo una cresta particular, se encuentra que avanza en el grupo y se hace gradualmente más débil y desaparece al frente, y mística y misteriosamente una débil de atrás se abre camino hacia adelante volviéndose más fuerte. En suma, las ondas se están moviendo a través del grupo mientras que éste sólo se mueve a la mitad de la velocidad que se mueven las ondas.

Como las velocidades de grupo y las de fase no son iguales, las ondas producidas por un objeto en movimiento ya no son un cono simplemente, sino algo mucho más interesante. Podemos ver esto en la figura 51-10, que muestra las ondas producidas por un objeto en movimiento sobre el agua. Observen que es completamente diferente de lo que tendríamos para el sonido, en el cual la velocidad es independiente de la longitud de onda, donde tendríamos frentes

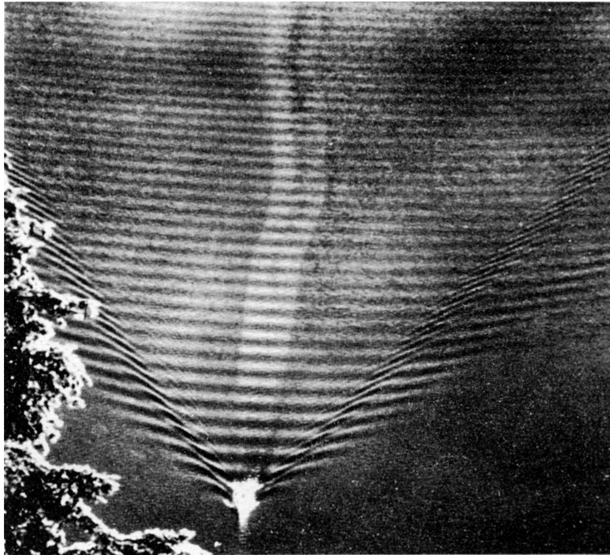


Figura 51-10. La estela de una lancha.

de onda únicamente a lo largo del cono viajando hacia afuera. En lugar de eso tenemos ondas atrás con frentes que se mueven paralelamente al movimiento de la lancha, luego tenemos onditas laterales a otros ángulos. Con ingenio, todo este diagrama de ondas se puede analizar conociendo únicamente esto: que la velocidad de fase es proporcional a la raíz cuadrada de la longitud de onda. El truco es que el diagrama de ondas es estacionario respecto a la lancha (a velocidad constante), cualquier otro diagrama se distanciaría de la lancha.

Las ondas de agua que hemos estado considerando hasta ahora eran ondas largas en las que la fuerza de restauración se debe a la gravedad. Sin embargo, cuando las ondas en el agua se acortan mucho, la principal fuerza de restauración es la acción capilar, es decir, la energía de la superficie, la tensión superficial. Para ondas de tensión superficial resulta que la velocidad de fase es

$$v_{\text{fase}} = \sqrt{2\pi T/\lambda\rho} \text{ (para los rizos)},$$

donde T es la tensión superficial y ρ la densidad. Es exactamente lo contrario: la velocidad de fase es *más alta* cuanto más corta es la longitud de onda, cuando la longitud de onda se hace muy pequeña. Cuando tenemos tanto la acción de la gravedad como la capilar, como ocurre siempre, obtenemos la combinación de estas dos:

$$v_{\text{fase}} = \sqrt{Tk/\rho + g/k},$$

donde $k = 2\pi/\lambda$ es el número de onda. Así, pues, la velocidad de las ondas de agua es bastante complicada. La figura 51-11 muestra la velocidad de fase en función de la longitud de onda. Para ondas muy cortas es grande, para ondas muy largas es grande, habiendo una velocidad mínima a la que las ondas pueden avanzar. Se puede calcular la velocidad de grupo a partir de la fórmula: es $\frac{2}{3}$ de la velocidad de fase para los rizos y $\frac{1}{2}$ de la velocidad de fase para las ondas gravitacionales. A la izquierda del mínimo la velocidad de grupo es más alta que la velocidad de fase. A la derecha, la velocidad de grupo es menor que la velocidad de fase. Hay una cantidad de fenómenos interesantes asociados con estos hechos. En primer lugar, como la velocidad de grupo aumenta tan rápidamente al disminuir la longitud de onda, si hacemos una perturbación habrá un extremo más lento de la perturbación yendo a la velocidad mínima con la longitud de onda correspondiente, y luego al frente, yendo a la velocidad más alta, habrá una onda corta y una onda muy larga. Es muy difícil ver las largas, pero es fácil ver las cortas en un tanque de agua.

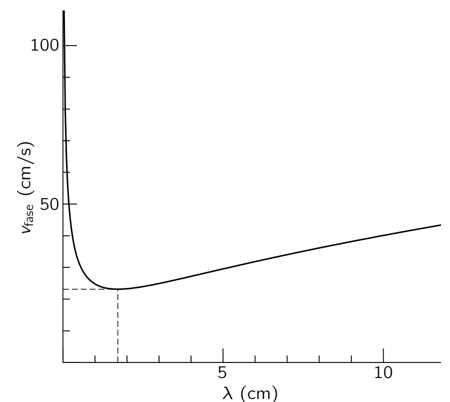


Figura 51-11. Velocidad de fase en función de la longitud de onda para el agua.

Vemos así que los rizos que se usan a menudo como ejemplo de ondas simples son completamente interesantes y complicados. No tienen de ninguna manera un

frente definido, como en el caso de las ondas simples como el sonido y la luz. La onda principal tiene pequeños rizos que corren hacia adelante. Una perturbación bien definida del agua no produce una onda bien definida debido a la dispersión. Primero vienen las ondas más finas. Entre paréntesis, si un objeto se mueve en el agua a cierta velocidad, resulta un diagrama bastante complicado porque las diversas ondas andan con velocidades diferentes. Se puede demostrar esto con una bandeja con agua y ver que las más rápidas son las ondas capilares finas. Hay ondas lentas, de cierto tipo, que marchan detrás. Inclinando el fondo se ve que donde la profundidad es menor, la velocidad es menor. Si entra una onda a cierto ángulo respecto a la línea de máxima pendiente, se tuerce y tiende a seguir esa línea. De este modo se puede demostrar diversas cosas y concluimos que las ondas son más complicadas en el agua que en el aire.

La velocidad de las ondas largas en agua con movimientos circulares es menor cuando la profundidad es menor, mayor en aguas profundas. Por lo tanto, cuando el agua avanza hacia una playa donde la profundidad disminuye, las ondas andan más lentamente. Pero donde el agua es más profunda, las ondas son más rápidas, por lo que obtenemos los efectos de ondas de choque. Aquí, como la onda no es tan simple, las ondas de choque son mucho más retorcidas, y la onda se repliega sobre sí misma en la forma familiar mostrada en la figura 51-12. Esto es lo que ocurre cuando las olas llegan a la playa, y la verdadera complejidad de la naturaleza se revela bien en esa circunstancia. Nadie ha podido todavía calcular qué forma debería tomar la ola al romper. Es bastante fácil cuando las olas son pequeñas, pero cuando una se agranda y rompe es mucho más complicado.



Figura 51-12. Onda de agua.

Se puede ver una característica interesante de las ondas capilares en las perturbaciones producidas por un objeto que se mueve en el agua. Desde el punto de vista del objeto mismo, el agua está pasando y las ondas que a la larga se establecen a su alrededor siempre son ondas que tienen justo la velocidad apropiada para permanecer quietas en el agua respecto al objeto. Análogamente, alrededor de un objeto en una corriente, con la corriente pasando, el dibujo de las ondas es estático y con las longitudes de onda justas para andar a la misma velocidad a que está pasando el agua. Pero si la velocidad de grupo es menor que la velocidad de fase, la perturbación se propaga hacia atrás por la corriente, porque la velocidad de grupo no es exactamente suficiente como para mantener el paso de la corriente. Si la velocidad de grupo es mayor que la velocidad de fase, el diagrama de ondas aparecerá al frente del objeto. Si se observa atentamente un objeto en una corriente, se puede ver que hay pequeños rizos al frente y largos "remolinos" atrás.

Otra característica interesante de esta clase se puede observar vertiendo líquidos. Por ejemplo, si se vierte leche de una botella lo bastante rápido, se

puede ver una gran cantidad de líneas atravesando en ambos sentidos la corriente que sale. Son ondas que parten de la perturbación en los bordes y se extienden en forma muy parecida a las ondas alrededor de un objeto en una corriente. Hay efectos provenientes de ambos lados que producen el dibujo cruzado.

Hemos investigado algunas propiedades interesantes de las ondas y las diversas complicaciones de cómo la velocidad de fase depende de la longitud de onda, la velocidad de las ondas en profundidad, etc., que producen los fenómenos realmente complejos, y por lo tanto interesantes, de la naturaleza.

52

Simetría en las leyes físicas

52-1 Operaciones de simetría

El tema de este capítulo es lo que podemos llamar *simetría en las leyes físicas*. Ya hemos discutido ciertos rasgos de simetría en las leyes físicas en conexión con el análisis vectorial (capítulo 11), la teoría de la relatividad (capítulo 16) y la rotación (capítulo 20).

¿Por qué nos debe preocupar la simetría? En primer lugar, la simetría es fascinante para la inteligencia humana y todo el mundo gusta de objetos o diagramas que son de algún modo simétricos. Es un hecho interesante el que la naturaleza nos ofrezca con frecuencia ciertos tipos de simetría en los objetos que encontramos en el mundo que nos rodea. Quizás el objeto más simétrico imaginable sea una esfera, y la naturaleza está llena de esferas —estrellas, planetas, gotitas de agua de las nubes—. Los cristales encontrados en las rocas presentan muchas clases diferentes de simetría, y el estudio de las mismas nos dice algunas cosas importantes acerca de la estructura de los sólidos. Aun los mundos animal y vegetal muestran algún grado de simetría, aunque la simetría de una flor o de una abeja no sea tan perfecta o tan fundamental como la de un cristal.

Pero nuestro principal interés aquí no es el hecho de que los *objetos* de la naturaleza son con frecuencia simétricos. Más bien, deseamos examinar algunas de las simetrías del universo que son aún más notables —las simetrías que existen en las mismas *leyes básicas* que gobiernan la marcha del mundo físico—.

Ante todo ¿qué *es* simetría? ¿Cómo puede una *ley* física ser "simétrica"? El problema de definir simetría es interesante y ya hemos apuntado que Weyl dio una buena definición, cuya esencia es: una cosa es simétrica si hay algo que podamos hacer con ella de tal modo que después que lo hemos hecho parece la misma cosa que antes. Por ejemplo, un jarrón simétrico es tal que reflejándolo o girándolo tendrá el mismo aspecto que antes. La cuestión que queremos considerar aquí es: qué podemos hacer a los fenómenos físicos o a una situación física en un experimento, y no alterar el resultado. La tabla 52-1 muestra una lista de operaciones conocidas frente a las cuales diversos fenómenos físicos permanecen invariantes.

Tabla 52-1

Operaciones de simetría

Traslación en el espacio
Traslación en el tiempo
Rotación en un ángulo fijo
Velocidad uniforme en línea recta (transformación de Lorentz)
Inversión del tiempo
Reflexión del espacio
Intercambio de átomos idénticos o partículas idénticas
Fase cuántica-mecánica
Materia-antimateria (conjugación de carga)

52-2 Simetría en el espacio y en el tiempo

La primera cosa que podríamos tratar de hacer, por ejemplo, es *trasladar* el fenómeno en el espacio. Si realizamos un experimento en una cierta región y luego construimos otro aparato en otro lugar del espacio (o trasladamos allí el original), cualquier cosa que pasó en un aparato, en un cierto orden de tiempo, ocurrirá del mismo modo si hemos arreglado la misma condición, con las atenciones debidas a las restricciones que mencionamos antes: que todas las características del ambiente que estorban para que se comporte igual se han quitado —hemos hablado de cómo definir cuánto deberíamos incluir en estas circunstancias y no entraremos en estos detalles de nuevo—.

Del mismo modo, también creemos hoy que el *desplazamiento en el tiempo* no produce ningún efecto en las leyes físicas. (Esto es, *en cuanto a lo que sabemos hoy en día* —todas estas cosas son así ¡en cuanto conocemos hoy en día!—) Esto significa que si construimos un cierto aparato y lo hacemos funcionar en un cierto tiempo, digamos el jueves a las 10 de la mañana, y luego construimos el mismo aparato y lo ponemos a funcionar, digamos, tres días más tarde en las mismas condiciones, los dos aparatos experimentarán los mismos movimientos

exactamente en la misma forma en función del tiempo, cualquiera que sea el instante de inicio de funcionamiento, teniendo presente de nuevo, naturalmente, que las características pertinentes del ambiente están también modificadas apropiadamente en el *tiempo*. Esta simetría significa, por supuesto, que si alguien compró las acciones de General Motors hace tres meses, ¡lo mismo le sucedería si las comprase ahora!

También tenemos que tener en cuenta las diferencias geográficas, ya que hay variaciones en las características de la superficie de la Tierra. Por ejemplo, si medimos el campo magnético en una cierta región y movemos el aparato a alguna otra región, puede ser que no trabaje precisamente del mismo modo ya que el campo magnético es diferente, pero decimos que esto ocurre porque el campo magnético está asociado con la Tierra. Podemos imaginar que si movemos el conjunto de la Tierra y el equipo, no habría variación en el funcionamiento del aparato.

Otra cosa que hemos discutido con mucho detalle fue la rotación en el espacio: si giramos un aparato un cierto ángulo, trabaja del mismo modo, siempre que también giremos con él todo lo que sea pertinente. De hecho, discutimos el problema de la simetría frente a una rotación en el espacio con algún detalle en el capítulo 11 e inventamos un sistema matemático llamado *análisis vectorial* para manejarlo lo más claramente posible.

En un nivel más avanzado nos encontramos con otra simetría —la simetría frente a velocidad uniforme en línea recta—. Es decir —un efecto bastante notable—, que si tenemos una pieza de un aparato trabajando de un cierto modo y tomamos el aparato y lo colocamos en un automóvil, y movemos todo el automóvil más lo pertinente del ambiente a una velocidad uniforme en línea recta, entonces en lo que respecta a los fenómenos dentro del automóvil no encontramos diferencia, todas las leyes físicas aparecen iguales. Y hasta sabemos cómo expresar esto más técnicamente: las ecuaciones matemáticas de las leyes físicas deben permanecer invariantes frente a una *transformación de Lorentz*. En realidad, fue un estudio del problema de la relatividad el que concentró con más interés la atención de los físicos en la simetría de las leyes físicas.

Ahora bien, todas las simetrías mencionadas anteriormente han sido de naturaleza geométrica, siendo más o menos iguales el espacio y el tiempo, pero hay otras simetrías de un tipo diferente. Por ejemplo, hay una simetría que describe el hecho de que podemos remplazar un átomo por otro de la misma clase; para decirlo de otro modo, *hay* átomos de la misma clase. Es posible encontrar grupos de átomos tales que si intercambiamos un par de ellos no ocurre nada —los átomos son idénticos—. Cualquier cosa que haga un átomo de oxígeno de un cierto tipo, hará otro átomo de oxígeno del mismo tipo. Alguien puede decir: “Esto es ridículo, ¡ésa es la *definición* de tipos iguales!” Puede que sea simplemente la definición, pero nosotros aún no sabemos si *existen* “átomos del mismo tipo”; la *realidad* es que hay muchos, muchos átomos del mismo tipo. Entonces, significa algo decir que no importa si remplazamos un átomo por otro del mismo tipo. Las llamadas partículas elementales que forman los átomos también son partículas idénticas en el sentido explicado —todos los electrones, los protones, los piones positivos, etc., son iguales—.

Después de una lista tan larga de cosas que se pueden hacer sin cambiar los fenómenos, uno podría pensar que se podría hacer cualquier cosa; demos algunos ejemplos en contra para ver la diferencia. Supongan que preguntamos: “¿Son las leyes físicas simétricas frente a un cambio de escala?” Supongan que construimos una cierta pieza del aparato y que luego construimos un aparato cinco veces mayor en cada una de sus partes, ¿trabajará exactamente del mismo modo? La respuesta, en este caso, es ¡*no*! La longitud de onda emitida, por ejemplo, por los átomos dentro de una caja de átomos de sodio y la longitud de onda de luz emitida por un gas de átomos de sodio con un volumen cinco veces mayor, no es cinco veces más larga, sino que es en realidad exactamente igual a la otra. Por lo que el cociente entre la longitud de onda y el tamaño del emisor cambiará.

Otro ejemplo: vemos en el periódico, de vez en cuando, fotos de alguna gran catedral construida con palillos de fósforos —una obra de arte fantástica realizada

por algún tipo jubilado que se entretiene encolando palillos de fósforos—. Es mucho más elaborada y maravillosa que cualquier catedral real. Si imaginamos que esta catedral de madera se construyese en realidad a escala de una catedral verdadera, vemos dónde se encuentra la dificultad; no duraría, el conjunto se desplomaría debido a que los palillos de fósforo construidos a escala no son lo suficientemente resistentes. "Sí", alguien podría decir, "pero también sabemos que cuando hay una influencia externa, también se debe cambiar proporcionalmente". Estamos hablando de la habilidad del objeto para resistir la gravitación. Por lo tanto, lo que deberíamos hacer en primer lugar es tomar la catedral modelo de fósforos reales y la tierra real, y entonces sabemos que es estable. Luego deberíamos tomar la catedral más grande y una tierra mayor. Pero entonces es aún peor, ¡porque la gravitación ha aumentado aún más!

Hoy en día, por supuesto, comprendemos el hecho de que los fenómenos dependen de la escala basándonos en que la materia es atómica por naturaleza y ciertamente si construyésemos un aparato que fuese tan pequeño que sólo contuviera cinco átomos, sería claramente algo que no podríamos hacer a una escala mayor o menor arbitrariamente. La escala de un átomo individual no es arbitraria de ningún modo —es algo definido—.

El hecho de que las leyes físicas no quedan invariantes frente a un cambio de escala fue descubierto por Galileo. Comprobó que las resistencias de los materiales no estaban exactamente en proporción justa a sus tamaños e ilustró esta propiedad que acabamos de discutir, acerca de la catedral de palillos de fósforos, dibujando dos huesos: el hueso de un perro en la proporción justa para sostener su peso, y el hueso imaginario de un "superperro" que sería, digamos, 10 o 100 veces mayor —dicho hueso era una cosa grande y sólida con proporciones muy diferentes—. No sabemos si llevó su argumento hasta la conclusión de que las leyes de la naturaleza deben tener una escala definida, pero se impresionó tanto con su descubrimiento que lo consideró tan importante como el descubrimiento de las leyes del movimiento, ya que los publicó en el mismo volumen, titulado "Sobre dos nuevas ciencias".

Otro ejemplo en el que las leyes no son simétricas, y que conocemos bastante bien, es éste: un sistema en rotación a velocidad angular uniforme no da las mismas leyes aparentes que uno que no rota. Si hacemos un experimento y luego lo colocamos en una nave espacial y la tenemos girando en el espacio vacío, completamente sola a una velocidad angular constante, el aparato no trabajará en la misma forma porque, como sabemos, las cosas dentro del equipo serán lanzadas hacia afuera, etc., debido a la fuerza centífruga o de coriolis, etc. De hecho, podemos decir que la Tierra está girando sin mirar fuera, si usamos un péndulo de Foucault.

Mencionaremos ahora una simetría muy interesante que evidentemente es falsa, es decir, *la reversibilidad en el tiempo*. Las leyes físicas aparentemente no pueden ser reversibles en el tiempo porque, como sabemos, todos los fenómenos evidentes son irreversibles en una escala grande: "El dedo que se mueve escribe, y habiendo escrito continúa". En cuanto a lo que podemos decir, esta irreversibilidad es debida al gran número de partículas involucradas y si pudiésemos ver las moléculas individuales no seríamos capaces de discernir si la maquinaria estaba trabajando hacia adelante o hacia atrás. Para ser más precisos: construimos un pequeño aparato en el que sabemos lo que están haciendo todos los átomos, en el que podemos observarlos agitándose. Después construimos otro aparato como el anterior, pero que empieza su movimiento en el estado final del otro, con todas las velocidades exactamente invertidas. *Entonces, ejecutará los mismos movimientos, pero exactamente al revés.* Dicho de otro modo: si tomamos una película, con detalle suficiente, de todos los mecanismos internos de un trozo de material y la proyectamos hacia atrás en una pantalla, ningún físico será capaz de decir: "esto es contra las leyes de la física, ¡esto está haciendo algo erróneo!" Si no vemos todos los detalles, naturalmente, la situación estará perfectamente clara. Si vemos un huevo estrellándose en la acera y la cáscara que se abre resquebrajándose, etc., entonces seguramente diremos: "esto es irreversible, ya que si proyectamos la película hacia atrás el huevo se volverá a formar y la cáscara

a ser entera, ¡y eso evidentemente es ridículo! Pero si consideramos a los átomos mismos individualmente, las leyes se presentan completamente reversibles. Éste es, naturalmente, un descubrimiento mucho más difícil de llevar a cabo, pero aparentemente es verdad que las leyes físicas fundamentales, a nivel microscópico y fundamental, ¡son completamente reversibles en el tiempo!

52-3 Simetría y leyes de conservación

Las simetrías de las leyes físicas son muy interesantes a este nivel, pero resulta, al fin, que son aún más interesantes y excitantes cuando llegamos a la mecánica cuántica. Por una razón que no podemos aclarar al nivel de la presente discusión —un hecho que la mayoría de los físicos aún encuentran algo desconcertante, una cosa muy profunda y bella—, es que, en mecánica cuántica, *para cada una de las reglas de simetría hay una ley de conservación correspondiente*, hay una conexión definida entre las leyes de conservación y las simetrías de las leyes físicas. Sólo podemos decir esto por ahora sin ningún intento de explicación.

Por ejemplo, el hecho de que las leyes son simétricas para la traslación en el espacio, cuando añadimos los principios de mecánica cuántica, resulta que significa que el *momentum se conserva.*

La invariancia frente a una rotación de un ángulo fijo en el espacio corresponde a la *conservación del momentum angular.* Estas conexiones son cosas muy interesantes y bellas, entre las cosas más bellas y profundas de la física.

Entre paréntesis, hay una cantidad de simetrías que aparecen en mecánica cuántica que no tienen análogo clásico, que no tienen método de descripción en física clásica. Una de ellas es la siguiente: si ψ es la amplitud de algún proceso, entonces el cuadrado del módulo de ψ es la probabilidad de que el proceso ocurra. Ahora bien, si algún otro tuviese que hacer sus cálculos, no con esta ψ, sino con una ψ' que difiere simplemente en un cambio de fase (sea Δ una constante, y multipliquemos $e^{i\Delta}$ por la vieja ψ), el cuadrado del módulo de ψ', que es la probabilidad del evento, es entonces igual al cuadrado del módulo de ψ:

$$\psi' = \psi e^{i\Delta}; \qquad |\psi'|^2 = |\psi|^2. \tag{52.1}$$

Por tanto, las leyes físicas no varían si la fase de la función de onda está corrida en una constante arbitraria. Ésta es otra simetría. Las leyes físicas deben ser de tal naturaleza que un corrimiento de la fase cuántica no altere nada. Como acabamos de decir, en mecánica cuántica hay una ley de conservación para cada simetría. La ley de conservación que se relaciona con la fase cuántica parece ser la *conservación de la carga eléctrica.* ¡Todo esto en conjunto es un negocio muy interesante!

52-4 Reflexiones en un espejo

La siguiente cuestión que nos va a ocupar por la mayor parte del resto de este capítulo es la cuestión de la simetría frente a *reflexión en el espacio.* El problema es éste: ¿son las leyes físicas simétricas frente a reflexión? Lo podemos decir así: supongan que construimos un equipo, digamos un reloj, con montones de ruedas y manecillas y números; suena, funciona y tiene cosas enrolladas dentro. Contemplamos el reloj en el espejo. La cuestión no es qué *aspecto* tiene en el espejo. En cambio, *construyamos* realmente otro reloj exactamente igual al aspecto que el primero tiene en el espejo; cada vez que hay un tornillo con cuerda derecha en uno, usamos un tornillo con cuerda izquierda en el lugar correspondiente del otro, donde uno tiene un "2" en la carátula, colocamos un "ᘔ" en la carátula del otro, cada resorte está enrollado en un sentido en uno y en el sentido opuesto en el reloj imagen. Cuando todo esté terminado, tenemos dos relojes, ambos físicos, que guardan entre sí la relación de un objeto con su imagen en el espejo, aunque ambos son, recalcamos, objetos materiales, reales. La pregunta ahora es: si los dos relojes comienzan en la misma condición, dándoles correspondientemente la misma cuerda, ¿harán tic-tac y girarán los dos relojes para siempre como imágenes en el espejo perfectas? (Ésta es una pregunta física

y no filosófica.) Nuestra intuición de las leyes físicas nos sugeriría que *deberían hacerlo*.

Sospecharíamos que, por lo menos en el caso de estos relojes, la reflexión en el espacio es una de las simetrías de las leyes físicas, que si lo cambiamos todo de "derecha" a "izquierda" y lo dejamos por lo demás igual, no podemos notar la diferencia. Supongamos, entonces, por un momento que esto es verdad. Si fuera verdad, entonces sería imposible distinguir "derecha" e "izquierda" mediante cualquier fenómeno físico; como es imposible, por ejemplo, definir una velocidad absoluta particular mediante un fenómeno físico. Por lo tanto, sería imposible definir absolutamente, mediante un fenómeno físico, lo que entendemos por "derecha" en oposición a "izquierda", puesto que las leyes físicas deberían ser simétricas.

Naturalmente, el mundo no *tiene* por qué ser simétrico. Por ejemplo, usando lo que podemos llamar "geografía", seguramente se puede definir "derecha". Supongan que estamos en Nueva Orleans y miramos hacia Chicago; Florida está a nuestra derecha (¡cuando tenemos los pies en el suelo!). Así, podemos definir "derecha" e "izquierda" mediante la geografía. Naturalmente que la situación real en cualquier sistema no tiene por qué tener la simetría de que estamos hablando, es una cuestión de si las *leyes* son simétricas —en otras palabras, si es *contra las leyes físicas* el tener una esfera como la Tierra con "cosas zurdas" y una persona como nosotros de pie mirando a una ciudad como Chicago desde un lugar como Nueva Orleans, pero con todo al revés de modo que Florida se encuentre al otro lado—. Claramente no parece imposible, ni contra las leyes físicas el que todo tenga derecha e izquierda intercambiadas.

Otro punto es que nuestra definición de "derecha" no debería depender de la historia. Un modo fácil de distinguir derecha de izquierda es ir a una ferretería y tomar un tornillo al azar. Las probabilidades son que tenga cuerda derecha —no necesariamente, pero es mucho más probable que la tenga a derecha que a izquierda—. Esto es una cuestión histórica o convencional, o del modo en que ocurre que son las cosas, y tampoco aquí es cuestión de leyes fundamentales. Como podemos apreciar bien, ¡todo el mundo podría haber comenzado a fabricar tornillos izquierdos!

De este modo, debemos tratar de encontrar algún fenómeno en el que "a derecha" sea una cosa fundamental. La próxima posibilidad que discutimos es el hecho de que luz polarizada rota su plano de polarización cuando atraviesa, digamos, agua azucarada. Además, como vimos en el capítulo 33, rota, digamos a la derecha, en una cierta solución de azúcar. La polarización va a la derecha si podemos disolver un poco de azúcar en agua, ésta será un manera de definir "a la derecha". Pero el azúcar proviene de cosas vivientes y si tratamos de fabricar azúcar artificialmente, descubrimos que ¡*no* rota el plano de polarización! Pero si tomamos esta misma azúcar que hemos fabricado artificialmente y que no rota el plano de polarización, y colocamos en ella bacterias (se comen un poco de azúcar) y luego filtramos las bacterias, encontramos que nos queda aún azúcar (casi la mitad de la que teníamos antes) y ahora sí rota el plano de polarización, ¡pero hacia *el otro lado*! Parece muy confuso, pero se ha explicado sencillamente.

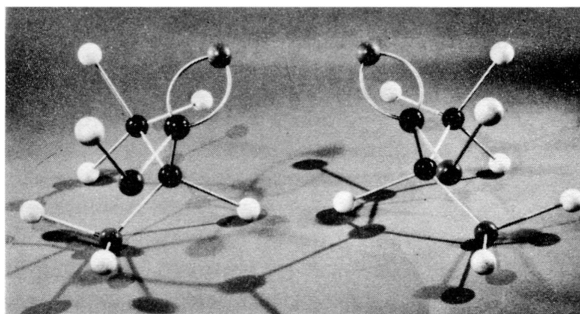


Figura 52-1. (a) L-alanina (izquierda) y (b) D-alanina (derecha).

Tomemos otro ejemplo: una de las sustancias que es común a todas las criaturas vivientes y fundamental para la vida es la proteína. Las proteínas consisten en cadenas de aminoácidos. La figura 52-1 muestra un modelo de aminoácido que resulta de una proteína. Este aminoácido se llama alanina, y el arreglo molecular se parecería al de la figura 52-1(a) si proviniera de una proteína de un ser viviente real. Por otro lado, si tratamos de hacer alanina a partir de dióxido de carbono, metano y amoniaco (y la *podemos* hacer, no es una molécula complicada), ¡descubrimos que estamos formando cantidades iguales de esta molécula y de la que muestra la figura 52-1(b)! La primera molécula, la que proviene de un ser viviente, se llama *L-alanina*. La otra, químicamente igual, porque tiene las mismas clases de átomos y las mismas conexiones entre ellos, es una molécula "a la derecha", comparada con la L-alanina "a la izquierda", y se llama *D-alanina*. Lo interesante es que cuando fabricamos alanina en casa en un laboratorio a partir de gases simples, obtenemos una mezcla de ambas por igual. Sin embargo, lo que la vida usa solamente es L-alanina. (Esto no es exactamente verdadero. Aquí y allá en las criaturas vivientes hay un uso especial de D-alanina, pero muy raramente. Todas las proteínas usan L-alanina exclusivamente.) Si fabricamos ambas clases y damos la mezcla a algún animal al que le guste "comer" o consumir alanina, no puede utilizar la D-alanina, por lo que solamente utiliza la L-alanina; esto es lo que sucedió con nuestra azúcar —después que las bacterias comen el azúcar que les conviene ¡solamente dejan la clase "equivocada"!— (el azúcar la izquierda sabe dulce, pero no igual que el azúcar la derecha).

Así parece como si los fenómenos de la vida permiten una distinción entre "derecha" e "izquierda", o la química permite una distinción ya que las dos moléculas son químicamente diferentes. Pero no, ¡no ocurre así! En tanto se puedan realizar medidas físicas, como de energía, velocidades de reacciones químicas, etc., las dos clases funcionan exactamente igual si realizamos todo también en una imagen en el espejo. Una molécula rotará la luz a la derecha, y la otra a la izquierda precisamente en la misma cantidad, si utilizamos la misma cantidad de fluido. Así, pues, en lo que respecta a la física, estos dos aminoácidos son igualmente satisfactorios. Como nosotros entendemos las cosas hoy en día, los fundamentos de la ecuación de Schrödinger exigen que las dos moléculas se deberían comportar de modos exactamente correspondientes, de manera que una es a derecha y otra a izquierda. No obstante, ¡en la vida todo ocurre de un solo modo!

Se presume que la razón de esto es la siguiente. Supongamos, que de alguna manera la vida está en un momento en una cierta condición, en la que todas las proteínas en algunas criaturas tienen aminoácidos a la izquierda, y que todas las enzimas son desequilibradas —toda sustancia en la criatura viviente es desequilibrada, no es simétrica—. Así, cuando las enzimas digestivas tratan de cambiar los compuestos químicos de la comida de una clase a otra, una clase de compuesto "le viene bien" a la enzima, pero no así el otro (como la Cenicienta y la zapatilla, excepto que es un "pie izquierdo" el que estamos probando). En cuanto a lo que sabemos, en principio, podríamos construir una rana, por ejemplo, en la que cada molécula estuviese invertida, todo es como la imagen en un espejo "izquierda" de una rana real, tenemos una rana izquierda. Esta rana izquierda andaría bien durante un momento, pero no encontraría nada que comer, porque si se traga una mosca sus enzimas no están construidas para digerirla. La mosca tiene la clase "equivocada" de aminoácidos (a menos que le demos una mosca a la izquierda). Por lo tanto, hasta donde nosotros sabemos, los procesos químicos y vitales continuarían del mismo modo si todo se invirtiese.

Si la vida es enteramente un fenómeno físico y químico, podemos entender que las proteínas están todas formadas con el mismo sacacorchos, solamente a partir de la idea de que al comienzo de todo algunas moléculas vivientes, accidentalmente, empezaron y unas pocas vencieron. En algún lugar, una vez, una molécula orgánica se desequilibró de una cierta manera, y de esta cosa particular sucedió que "derecha" empezó a evolucionar en nuestra geografía particular, un accidente histórico determinado fue unilateral, y por siempre desde entonces el desequilibrio se ha propagado. Una vez que ha llegado al estado en

que se encuentra ahora, naturalmente, continuará siempre —todas las enzimas digieren las cosas a la derecha, fabrican las cosas a la derecha—. En el momento en que el dióxido de carbono y el vapor de agua, etc., llegan a las hojas de las plantas, las enzimas que hacen el azúcar los desequilibran porque ellas están desequilibradas. Si cualquier clase nueva de virus o cosa viviente se originase en un tiempo posterior, sobreviviría solamente si pudiese "comer" la clase de materia viviente ya presente. Por lo tanto, él también debe ser de la misma clase.

No hay conservación del número de moléculas a la derecha. Una vez empezado, podríamos seguir aumentando el número de moléculas a la derecha. Por consiguiente, la suposición es, entonces, que en el caso de la vida los fenómenos no demuestran una falta de simetría en las leyes físicas, pero sí demuestran, por el contrario, la naturaleza universal y lo común del origen último de todas las criaturas de la tierra, en el sentido anteriormente descrito.

52-5 Vectores polares y axiales

Ahora avanzamos más. Observamos que hay una gran cantidad de lugares en física donde tenemos reglas "de la mano derecha" y "de la mano izquierda". De hecho, al estudiar análisis vectorial, aprendimos las reglas de la mano derecha que teníamos que utilizar para poder obtener el momentum angular, el torque, el campo magnético, etcétera. Por ejemplo, la fuerza sobre una carga que se mueve en un campo magnético es $\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}$. En una situación dada, en la que conocemos $\boldsymbol{F}$, $\boldsymbol{v}$ y $\boldsymbol{B}$, ¿no es esta ecuación suficiente para definir la derecha? En realidad, si volvemos atrás y miramos de dónde provenían los vectores, sabemos que la "regla de la mano derecha" era simplemente una convención, era un truco. Las magnitudes originales, como los momenta angulares y las velocidades angulares, y otras cosas de esta especie, ¡no eran realmente vectores! Todos ellos están de algún modo asociados con un cierto plano, y es solamente porque hay tres dimensiones en el espacio por lo que podemos asociar la cantidad con una dirección perpendicular a aquel plano. De las dos direcciones posibles, escogemos la dirección "mano derecha".

Por tanto, si las leyes de la física son simétricas, encontraríamos que si se escurriese algún demonio dentro de todos los laboratorios y remplazase la palabra "derecha" por "izquierda" en todos los libros donde se dan las "reglas de la mano derecha", y en su lugar tuviésemos que usar "reglas de la mano izquierda" uniformemente, esto no implicaría ninguna diferencia en las leyes físicas.

Demos un ejemplo. Hay dos tipos de vectores. Hay vectores "honestos", por ejemplo, un paso $\boldsymbol{r}$ en el espacio. Si en nuestro aparato hay aquí una pieza y allá algo más, entonces en el aparato espejo estará la pieza y el algo más, y si dibujamos un vector desde la "pieza" al "algo más", un vector está en la imagen del espejo (figura 52-2). La flecha del vector cambia de dirección, como si todo el espacio se volteara; este vector se llama *vector polar*.



Figura 52-2. Un paso en el espacio y su imagen especular.

Pero el otro tipo de vector, que tiene que ver con las rotaciones, es de naturaleza diferente. Por ejemplo, supongan que algo está rotando en tres dimensiones, como muestra la figura 52-3. Entonces, si lo miramos en un espejo, estará rotando como se indica, es decir, como la imagen en el espejo de la rotación original. Ahora bien, estamos de acuerdo en representar la rotación en la imagen del espejo mediante la misma regla: es un "vector" que, en la reflexión, *no* cambia como ocurre con el vector polar, sino que está invertido con relación a los vectores polares y a la geometría del espacio. Tal vector se llama *vector axial*.

Ahora bien, si la ley de simetría por reflexión es correcta en física, debe ser verdad que las ecuaciones se deben expresar de tal modo que si cambiamos el signo de cada vector axial y de cada producto vectorial, que sería lo que corresponde a la reflexión, nada sucederá. Por ejemplo, cuando escribimos una fórmula que dice que el momentum angular es $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$, esta ecuación es muy correcta, porque si cambiamos a un sistema de coordenadas a izquierda, cambiamos el signo de $\boldsymbol{L}$, pero $\boldsymbol{p}$ y $\boldsymbol{r}$ no cambian; se ha cambiado el signo del producto vectorial, puesto que debemos cambiar de una regla derecha a una izquierda. Otro ejemplo: sabemos que la fuerza sobre una carga que se mueve en un campo magnético es

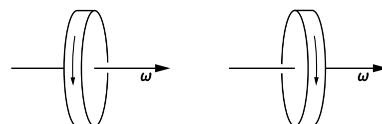


Figura 52-3. Una rueda que gira y su imagen en el espejo. Observen que el "vector" velocidad angular no ha invertido su dirección.

$\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}$, pero si cambiamos de un sistema derecho a uno izquierdo, como sabemos que $\boldsymbol{F}$ y $\boldsymbol{v}$ son vectores polares, el cambio de signo requerido por el producto vectorial se debe cancelar por un cambio de signo en $\boldsymbol{B}$, lo que significa que $\boldsymbol{B}$ debe ser un vector axial. En otras palabras, si efectuamos esa reflexión, $\boldsymbol{B}$ debe pasar a $-\boldsymbol{B}$. Por lo que si cambiamos nuestras coordenadas derecha por izquierda, también debemos intercambiar los polos norte y sur de los imanes.

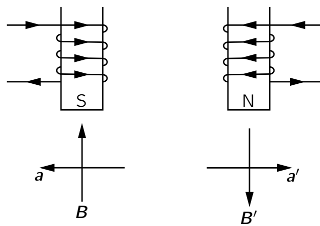


Figura 52-4. Un imán y su imagen especular.

Veamos cómo funciona esto en un ejemplo. Supongan que tenemos dos imanes, como en la figura 52-4. Uno es un imán con el bobinado en un sentido y corriente en una dirección dada. El otro imán parece la reflexión del primero en un espejo —el bobinado irá en el otro sentido, todo lo que sucede dentro del alambre es exactamente a la inversa y la corriente va como se muestra—. Ahora bien, a partir de las leyes para la producción da campos magnéticos, que no conocemos aún oficialmente, pero que aprendimos muy probablemente en la escuela secundaria, resulta que el campo magnético es como el que muestra la figura. En un caso el polo es un polo magnético sur, mientras que en el otro imán la corriente va en el otro sentido y el campo magnético está invertido —es un polo magnético norte—. Así, vemos que cuando intercambiamos derecha e izquierda, ¡debemos intercambiar norte y sur!

No les importa cambiar el norte al sur, también éstos son meras convenciones. Hablemos de *fenómenos*. Supongan ahora que tenemos un electrón moviéndose en un campo que entra en la página. Si usamos entonces la fórmula para la fuerza, $\boldsymbol{v} \times \boldsymbol{B}$ (recuerden que la carga es negativa), encontramos que el electrón se desviará en la dirección indicada de acuerdo con la ley física. Por lo tanto, el fenómeno es que tenemos un arrollamiento con corriente en un sentido específico y un electrón que se desvía en un cierto camino —esto es la física—, no importa cómo llamemos a cada cosa.

Hagamos ahora el mismo experimento con un espejo: enviamos un electrón en una dirección correspondiente y ahora la fuerza está invertida si la calculamos siguiendo la misma regla, y esto está muy bueno ¡porque los *movimientos* correspondientes son entonces imágenes en espejo!

52-6 ¿Cuál mano es la derecha?

Así, pues, la realidad de las cosas es que al estudiar cualquier fenómeno hay siempre dos reglas de la mano derecha, o un número par de ellas, y el resultado final es que el fenómeno siempre parece simétrico. En resumen, no podemos distinguir, por lo tanto, derecha de izquierda si tampoco somos capaces de distinguir norte de sur. Sin embargo, puede parecer que *podemos* decir cuál es el polo norte de un imán. El polo norte de la aguja de una brújula, por ejemplo, es el que apunta al norte. Pero naturalmente esto es de nuevo una propiedad local que tiene que ver con la geografía de la Tierra, esto es lo mismo que hablar de la dirección en que se encuentra Chicago, por lo que no cuenta. Si hemos visto agujas de brújulas, puede que hayamos notado que el polo que mira al norte es de una especie de color azulado. Pero se debe justamente al hombre que pintó el imán. Todos estos criterios son locales y convencionales.

Sin embargo, si un imán tuviese la propiedad de que si lo mirásemos muy de cerca, viésemos pelitos saliendo de su polo norte y no en su polo sur, si ésta fuese la regla general, o si hubiere *algún* modo único de distinguir el polo norte del sur en un imán, entonces podríamos decir cuál de los dos casos tendríamos realmente, *y esto sería el fin de la ley de simetría por reflexión.*

Para ilustrar aún más claramente todo el problema, imaginen que estuviésemos hablando a un marciano por teléfono, o a quienquiera que sea, pero muy lejos. No se nos permite enviarle ninguna muestra real para que la inspeccione. Por ejemplo, si pudiésemos enviar luz, podríamos enviarle luz polarizada circularmente a la derecha y decir: “esto es luz a la derecha, observe simplemente el modo en que llega”. Pero no podemos *darle* nada, solamente podemos hablarle. Está muy lejos, o en alguna ubicación extraña, y no puede ver nada de lo que nosotros vemos. Por ejemplo, no podemos decir: “Mire a la Osa Mayor, ahora mire cómo

están dispuestas estas estrellas. Lo que nosotros entendemos por 'derecha' es..."
Solamente se nos permite hablar por teléfono.

Ahora bien, nosotros queremos contarle todo lo nuestro. Naturalmente, comenzamos definiéndole en primer lugar los números y decimos: "Tic, tic, *dos*; tic, tic, tic, *tres*;...", de modo que gradualmente son capaces de entender un par de palabras, etc. Después de algún tiempo puede ser que nos hagamos muy amigos de ese tipo y él nos diga: "¿qué aspecto tienen ustedes?" Comenzamos a describirnos y decimos: "bueno, tenemos un metro setenta de estatura". Él nos corta: "un momento, ¿qué es un metro setenta?" ¿Es posible explicarle lo que es un metro setenta? ¡Ciertamente! Decimos: "usted conoce el diámetro de los átomos de hidrógeno, ¡nosotros tenemos 17 000 millones de átomos de hidrógeno de altura!" Esto es posible ya que las leyes físicas son invariantes frente a un cambio de escala y, por lo tanto, *podemos* definir una longitud absoluta. Y del mismo modo definimos el tamaño del cuerpo y le decimos cuál es el aspecto general, tiene dientes con cinco protuberancias que sobresalen en los extremos, etc., y nos sigue, y terminamos describiéndole nuestro aspecto externo, presumiblemente sin encontrar dificultades mayores. Es más, está haciendo un modelo nuestro a medida que avanzamos. Nos dice: "oye, ustedes son ciertamente tipos muy atractivos, pero, ¿qué tienen dentro? Comenzamos a describirle los diversos órganos internos, y llegamos al corazón y le describimos cuidadosamente su forma y decimos: "ahora coloca el corazón en el lado izquierdo". Nos responde: "¡queeeeé!, ¿el lado izquierdo?" Nuestro problema ahora es describirle de qué lado va el corazón sin que él vea nada de lo que nosotros vemos y sin enviarle jamás una muestra de lo que entendemos por "derecha" —ningún objeto patrón de derecha—. ¿Podemos hacer eso?

52-7 ¡La paridad no se conserva!

Resulta que las leyes de gravitación, las leyes de electricidad y magnetismo, las fuerzas nucleares, todas satisfacen el principio de la simetría por reflexión, por lo que no se pueden usar estas leyes, o cualquier cosa que se derive de ellas. Pero hay un fenómeno llamado *desintegración beta*, o *desintegración débil*, asociado con las muchas partículas que se han descubierto en la naturaleza. Un ejemplo de desintegración débil, en conexión con una partícula descubierta alrededor de 1954, originó un enigma extraño. Había cierta partícula cargada que se desintegraba en tres mesones π, como muestra esquemáticamente la figura 52-5. A esta partícula se le llamó durante algún tiempo mesón τ. Ahora bien, vemos también otra partícula, en la figura 52-5, que se desintegra en *dos* mesones; uno debe ser neutro, debido a la conservación de la carga. A esta partícula se le llamó mesón θ. Por lo que tenemos una partícula llamada τ que se desintegra en tres mesones π, y una partícula θ que se desintegra en dos mesones π. Pero pronto se descubrió que τ y θ tienen una masa casi igual, de hecho son iguales, dentro del error experimental. Luego se encontró que el tiempo que tardaban en desintegrarse en tres π y en dos π era casi exactamente el mismo, viven el mismo periodo de tiempo. Después que siempre que se formaban lo hacían en la misma proporción, digamos 14 % de τ y 86 % de θ.

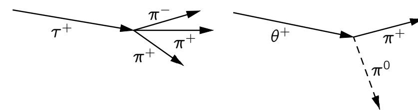


Figura 52-5. Diagrama esquemático de la desintegración de una partícula τ^+ y de una partícula θ^+.

Cualquiera con sentido cabal se da cuenta inmediatamente que deben ser la misma partícula, que producimos simplemente un objeto que tiene dos modos diferentes de desintegrarse, no dos partículas diferentes. Este objeto que se puede desintegrar de dos modos diferentes tiene, en consecuencia, el mismo tiempo de vida y la misma probabilidad de producción (ya que es sencillamente el cociente de las probabilidades con las cuales se desintegra en esas dos clases).

Sin embargo, fue posible probar (y aquí de ningún modo podemos explicar *cómo*), a partir del principio de reflexión en mecánica cuántica, que era *imposible* que los dos provengan de la misma partícula —la misma partícula *no podía* desintegrarse de ambos modos—. La ley de conservación correspondiente al principio de la simetría por reflexión es algo que no tiene análogo clásico y por lo tanto esta clase de conservación cuántica se llamó *conservación de la paridad*. Así, fue un resultado de la conservación de la paridad o, más precisamente, de la

simetría de las ecuaciones cuánticas para la desintegración débil frente a reflexión, el que la misma partícula no pudiese ir en las dos direcciones, por lo que debía ser una especie de coincidencia de masas, tiempos de vida, etc. Pero cuanto más se estudiaba, más notable era la coincidencia y gradualmente surgió la sospecha de que posiblemente la profunda ley de la simetría de la naturaleza respecto a reflexión pudiese ser falsa.

Como resultado de este fracaso aparente, los físicos Lee y Yang sugirieron que se deberían hacer otros experimentos en desintegraciones relacionadas para tratar de probar si la ley era correcta en otros casos. El primer experimento de esta clase lo hizo Miss Wu de Columbia y es como sigue: usando un imán muy potente a una temperatura muy baja, resulta que un cierto isótopo de cobalto que se desintegra emitiendo un electrón, es magnético, y si la temperatura es lo bastante baja para que las oscilaciones térmicas no sacudan demasiado los imanes atómicos, éstos se alinean en el campo magnético. Así, pues, los átomos de cobalto se alinearán en este campo intenso. Luego se desintegran emitiendo un electrón y se descubrió que cuando los átomos estaban alineados en un campo cuyo vector $\boldsymbol{B}$ apunta hacia arriba, la mayoría de los electrones se emitían hacia abajo.

Si no se está al tanto de los últimos adelantos del mundo, tal observación no parece nada significativa, pero si se aprecian los problemas y cosas interesantes del mundo, entonces se ve que es el descubrimiento más sorprendente, cuando colocamos átomos de cobalto en un campo magnético extremadamente intenso, van más electrones de desintegración hacia abajo que hacia arriba. Por tanto, si lo pusiéramos en un experimento correspondiente en un “espejo”, en el que los átomos de cobalto estarían alineados en la dirección opuesta, expulsarían sus electrones *hacia arriba* y no *hacia abajo*; la acción es *asimétrica. ¡Le han crecido pelos al imán!* El polo sur de un imán es de tal suerte que los electrones de una desintegración β tienden a alejarse de él, esto distingue de un modo físico, el polo norte del sur.

Después de éste se hicieron muchos otros experimentos: la desintegración de π en μ y ν, μ en un electrón y dos neutrinos; hoy en día Λ en un protón y π; desintegración de las Σ, y muchas otras desintegraciones. De hecho, en casi todos los casos donde se podía esperar ¡se ha encontrado que *ninguno* de ellos obedece la simetría de reflexión! La ley de la simetría de reflexión es incorrecta, fundamentalmente a este nivel de la física.

En resumen, podemos decirle a un marciano dónde colocar el corazón. Decimos: “escuche, construya un imán, póngale las bobinas y haga pasar corriente, tome entonces un poco de cobalto y baje la temperatura. Arregle el experimento de modo que los electrones marchen de sus pies a su cabeza; entonces, la dirección en la que la corriente recorre las bobinas es la dirección que entra en lo que nosotros llamamos la derecha y sale a la izquierda”. En consecuencia, es posible definir derecha e izquierda, ahora, haciendo un experimento de esta clase.

Se predijeron también muchas otras características. Por ejemplo, resulta que el espín, el momentum angular, del núcleo de cobalto antes de la desintegración es 5 unidades $\hbar$ y después de la desintegración es 4. El electrón lleva momentum angular de espín y también interviene un neutrino. Es fácil ver a partir de esto que el electrón debe llevar su momentum angular de espín alineado según su dirección de movimiento, y asimismo el neutrino. De este modo, parece como si el electrón estuviese girando a la izquierda y esto también se comprobó. De hecho, se comprobó precisamente aquí en Caltech por Boehm y Wapstra, que el electrón gira principalmente a la izquierda. (Hubo otros experimentos que dieron la respuesta opuesta, ¡pero estaban equivocados!)

El problema siguiente, por supuesto, fue a encontrar la ley del fracaso de la conservación de la paridad. ¿Cuál es la regla que nos dice cómo va a ser de grande el fracaso? Ésta: ocurre solamente en estas reacciones muy lentas, llamadas desintegraciones débiles, y cuando ocurre, la regla es que las partículas que tienen espín, como el electrón, el neutrino, etc., salen con un espín que tiende a la izquierda. Es una regla desequilibrada, relaciona un vector polar velocidad y

un vector axial momentum angular, y dice que es más probable que el momentum angular sea opuesto a la velocidad que a lo largo de ella.

Ésta es, pues, la regla, pero hoy no entendemos realmente los por qué y los por consiguiente de ella. *¿Por qué* es ésta la regla correcta, cuál es la razón fundamental de ello, y cómo se relaciona con cualquier otra cosa? Por el momento quedamos tan sorprendidos por el hecho de que es asimétrica, que no nos hemos podido recobrar lo suficiente para entender lo que significa respecto a todas las otras reglas. Sin embargo, es un asunto interesante, moderno y aún sin resolver, por lo que parece apropiado discutir algunas cuestiones asociadas con él.

52-8 Antimateria

Lo primero que hay que hacer cuando se pierde una de las simetrías es volver inmediatamente a la lista de simetrías conocidas o supuestas y preguntar si se pierde alguna de las otras. Ahora bien, no mencionamos una operación en nuestra lista, que debe ser necesariamente investigada, y es la relación entre materia y antimateria. Dirac predijo que además de los electrones debe haber otra partícula, llamada positrón (descubierta en Caltech por Anderson), que se relaciona necesariamente con el electrón. Todas las propiedades de estas dos partículas obedecen ciertas reglas de correspondencia: las energías son iguales; las masas son iguales; las cargas son opuestas; sin embargo, lo más importante de todo, las dos cuando se juntan se pueden aniquilar entre sí y liberar su masa completa en forma de energía, digamos rayos γ. El positrón se llama *antipartícula* del electrón, y ésas son las características de una partícula y su antipartícula. Resultó claro del razonamiento de Dirac que todas las demás partículas del mundo deberían tener también antipartículas correspondientes. Por ejemplo, para el protón debería haber un antiprotón, que se simboliza ahora con $\bar{p}$. El $\bar{p}$ tendría una carga eléctrica negativa, la misma masa que un protón, etc. Sin embargo, el rasgo más importante es que cuando se unen un protón y un antiprotón se pueden aniquilar entre sí. La razón por la que recalcamos esto es porque la gente no lo entiende cuando decimos que hay un neutrón y también un antineutrón, ya que dicen: "Un neutrón es neutro; por lo tanto, ¿cómo *puede* tener una carga opuesta?" La regla del "anti" no es solamente que tiene la carga opuesta, tiene cierto conjunto de propiedades y el conjunto total de las mismas es opuesto. El antineutrón se distingue del neutrón del siguiente modo: si colocamos dos neutrones juntos, permanecen simplemente como dos neutrones, pero si colocamos un neutrón y un antineutrón juntos, se aniquilan liberándose una gran explosión de energía, con diversos mesones π, rayos γ y muchas cosas más.

Ahora bien, si tenemos antineutrones, antiprotones y antielectrones, en principio podemos hacer antiátomos. No se han hecho aún, pero es posible en principio. Por ejemplo, un átomo de hidrógeno tiene un protón en el centro con un electrón dando vueltas. Ahora bien, imaginen que en algún lugar podemos hacer un antiprotón con un positrón dando vueltas alrededor, ¿daría vueltas? Bien, en primer lugar, el antiprotón es eléctricamente negativo y el antielectrón es eléctricamente positivo, por lo que se atraen uno al otro de una manera correspondiente —las masas son iguales, todo es igual—. Es uno de los principios de la simetría de la física —las ecuaciones parecen demostrarlo— si un reloj, digamos, se hiciese de materia por un lado, y luego hiciésemos el mismo reloj de antimateria, funcionaría de este modo. (Naturalmente, si colocamos los dos relojes juntos, se aniquilarían, pero esto es diferente.)

Entonces surge una cuestión inmediata. Podemos construir dos relojes de materia, uno que es "izquierdo" y uno que es "derecho". Por ejemplo, podríamos construir un reloj que no se construye de un modo simple, sino que tiene cobalto e imanes y detectores de electrones que detectan la presencia de electrones de desintegración β y los cuentan. Cada vez que se cuenta uno, el segundero se mueve. Entonces el reloj imagen, que recibe menos electrones, no irá a la misma velocidad. Pero evidentemente podemos construir dos relojes en forma tal que el reloj izquierdo no concuerde con el derecho. Construyamos de materia un reloj que llamaremos patrón o reloj derecho. Ahora construyamos, también de

materia, un reloj que llamaremos reloj izquierdo. Acabamos de descubrir que, en general, los dos *no* funcionarán del mismo modo. Anteriormente a este famoso descubrimiento físico se pensaba que lo harían. Pero también se supuso que la materia y la antimateria eran equivalentes. Esto es, si construyésemos un reloj de antimateria, derecho, de la misma forma, entonces funcionaría lo mismo que el reloj de materia derecho, y si hiciésemos el mismo reloj izquierdo funcionaría igual. En otras palabras, al principio se creyó que *todos* estos *cuatro* relojes eran iguales; ahora naturalmente sabemos que los de materia derecho e izquierdo no son iguales. En consecuencia, es presumible que los de antimateria derecho e izquierdo no son iguales.

Por lo tanto, la pregunta es: ¿cuál va con cuál, si es que lo hacen? En otras palabras, ¿se comporta el de materia derecho como el de antimateria derecho? ¿o es que el de materia derecho se comporta como el de antimateria izquierdo? Los experimentos de desintegración β que usan desintegración en positrones en lugar de desintegración en electrones, indican que ésta es la interconexión: la materia “derecha” funciona lo mismo que la antimateria “izquierda”.

En consecuencia, por fin, ¡es realmente verdad que la simetría derecha e izquierda se mantiene aún! Si hiciésemos un reloj izquierdo, pero hecho de la otra clase de materia, antimateria en lugar de materia, funcionaría igual. Así, lo que ha sucedido es que en lugar de tener dos reglas independientes en nuestra lista de asimetrías, dos de estas reglas se unen para formar una nueva regla: la materia a la derecha es simétrica de la antimateria a la izquierda.

Por lo que si nuestro marciano está formado de antimateria y le damos instrucciones para construir aquel modelo “a la derecha” como nosotros, saldrá, naturalmente, al revés. ¿Qué sucedería cuando, después de mucha plática de ambos lados, nos hemos enseñado el uno al otro a construir naves espaciales y nos encontramos a mitad de camino en el espacio vacío? Nos hemos aleccionado uno al otro acerca de nuestras costumbres, etc., y los dos nos apresuramos a estrecharnos las manos. Bien, si él ofrece su mano izquierda, ¡tenga cuidado!

52-9 Simetrías rotas

La pregunta siguiente es: ¿qué podemos construir con las leyes que son *casi* simétricas? Lo maravilloso de todo esto es que para una gama tan amplia de fenómenos intensos e importantes —fuerzas nucleares, fenómenos eléctricos, y aun los débiles como la gravitación— en una enorme variedad de campos de la física, todas sus leyes parecen ser simétricas. Por otro lado, este pequeño trozo adicional dice: “No, ¡las leyes no son simétricas!” ¿Cómo es que la naturaleza puede ser casi simétrica y no perfectamente simétrica? ¿Qué hacemos con esto? En primer lugar, ¿tenemos algún otro ejemplo? La respuesta es que de hecho sí tenemos unos pocos ejemplos más. Por ejemplo, la parte nuclear de la fuerza entre protón y protón, entre neutrón y neutrón, y entre neutrón y protón, es exactamente igual —hay una nueva simetría para las fuerzas nucleares, que podemos intercambiar neutrón y protón—, pero evidentemente no es una simetría general, ya que la repulsión eléctrica entre dos protones a distancia no existe para neutrones. Por lo que no es generalmente verdad que *siempre* podemos remplazar un protón por un neutrón, es sólo una buena aproximación. ¿Por qué *buena*? Porque las fuerzas nucleares son mucho más intensas que las fuerzas eléctricas. Así, ésta es también una “casi” simetría. Tenemos entonces ejemplos en otras cosas.

Tendemos, en nuestro interior, a aceptar la simetría como cierta clase de perfección. De hecho es como la vieja idea de los griegos de que las circunferencias eran perfectas y fue bastante horrible creer que las órbitas planetarias no eran circunferencias, sino sólo casi circunferencias. La diferencia entre ser una circunferencia y una casi circunferencia, no es una diferencia pequeña, es un cambio fundamental en lo que respecta a la inteligencia. En una circunferencia hay un signo de perfección y simetría que no se encuentra en el momento en que la circunferencia varía ligeramente —es el fin—, ya no es más simétrica. La cuestión es entonces por qué es solamente *casi* una circunferencia —ésta es una pregunta mucho más difícil—. El movimiento real de los planetas, en general,

debería ser elipses, pero a través de las edades, a causa de las fuerzas de marea, etc., se han transformado en casi simétricas. La cuestión es ahora ver si tenemos aquí un caso parecido. El problema desde el punto de vista de las circunferencias es que si fuesen circunferencias perfectas no habría nada que explicar, eso es claramente sencillo. Pero puesto que son solamente casi circunferencias, hay mucho que explicar, y el resultado implicó un gran problema dinámico y ahora nuestro problema es explicar por qué son casi simétricas considerando las fuerzas de marea, etcétera.

Así, pues, nuestro problema es explicar de dónde proviene la simetría. ¿Por qué la naturaleza es tan casi simétrica? Nadie tiene una idea de por qué. Lo único que podemos sugerir es algo como esto: hay una puerta en Japón, en Neiko, que los japoneses llaman algunas veces la puerta más hermosa de todo Japón, se construyó en una época en que había una gran influencia del arte chino. Esta puerta es muy ornamentada, con muchos tímpanos y hermosas estatuas y columnas y cabezas de dragones y príncipes esculpidos en los pilares, etc. Pero cuando se mira de cerca se ve que en el dibujo elaborado y complejo de uno de los pilares, uno de los pequeños elementos de diseño está esculpido cabeza abajo; por lo demás, todo es completamente simétrico. Si se pregunta por qué es así, la historia es que se esculpió cabeza abajo para que los dioses no estuviesen celosos de la perfección humana. Así, a propósito, cometieron un error en ella, para que los dioses no estuviesen celosos y se enfadasen con los seres humanos.

Nos gustaría invertir la idea y pensar que la verdadera explicación de la casi simetría de la naturaleza es ésta: ¡Dios hizo las leyes solamente casi simétricas para que así nosotros no estuviésemos celosos de Su perfección!

Índice analítico

E

F

G

H

I

J

L

S

T

V

W

Índice de nombres

T

V

W

Y

Z

Lista de símbolos

α	aceleración angular. 18-3
g	aceleración de la gravedad. 9-4
$\approx$	aproximadamente. 6-7
A	área. 5-9
L	auto-inductancia. 23-6
$P(k,n)$	Bernoulli o probabilidad binomial. 6-5
Q	calor. 44-3
C_V	calor específico a volumen constante. 45-2
C	capacitancia. 23-5
q	carga eléctrica. 12-8
$\binom{n}{k}$	coeficiente binomial, n sobre k. 6-4
μ	coeficiente de fricción. 12-4
a^*	complejo conjugado de a. 23-1
a_x, a_y, a_z	componentes cartesiana del vector de aceleración. 8-10
E_x, E_y, E_z	componentes cartesianas del vector de campo eléctrico. 12-11
B_x, B_y, B_z	componentes cartesianas del vector de campo magnético. 12-11
F_x, F_y, F_z	componentes cartesianas del vector de fuerza. 9-3
p_x, p_y, p_z	componentes cartesianas del vector de momentum. 10-9
k_x, k_y, k_z	componentes cartesianas del vector de onda. 34-10
v_x, v_y, v_z	componentes cartesianas del vector de velocidad. 8-10
κ	conductividad térmica. 43-11
k	constante de Boltzmann, $k = 1.3806504 \times 10^{-23}$ J/K. 39-10
h	constante de Planck, $h = 6.62606896 \times 10^{-34}$ Js. 17-9
$\bar{h}$	constante de Planck reducida, $\bar{h} = h/2\pi$. 2-7
σ	constante de Stefan-Boltzmann, $\sigma = 5.6704 \times 10^{-8} \mathrm{W/m^2K^2}$. 45-8
ϵ_0	constante dieléctrica o permitividad del vacío, $\epsilon_0 = 8.854187817 \times 10^{-12}$. 12-8
G	constante gravitacional. 7-9
x	coordenada cartesiana. 1-6
y	coordenada cartesiana. 1-6
z	coordenada cartesiana. 1-6
I	corriente eléctrica. 23-6
$\sim$	del orden. 2-11
ρ	densidad. 47-4
d	distancia. 12-7
s	distancia. 8-1
f'	distancia focal. 27-3
E	energía. 4-7

$\mathcal{E}$	energía. 33-11
T	energía cinética. 13-1
U	energía interna. 39-5, 39-6
U	energía potencial. 13-1
S	entropía. 44-11
ν	frecuencia. 17-9
$\hat{Z}$	impedancia compleja. 23-7
n	índice de refracción. 26-4
I	intensidad. 30-1
λ	longitud de onda. 17-9
a	magnitud o componente de vector de aceleración. 8-9
p	magnitud o componente de vector de momentum. 2-7
F	magnitud o componente del vector de fuerza. 7-1
L	magnitud o componente del vector de momento angular. 18-6
k	magnitud o componente del vector de onda, número de onda. 29-3
v	magnitud o componente del vector de velocidad. 8-5
m	masa. 4-7
m_0	masa de descanso. 10-9
I	momento de inercia. 18-8
$p_\mu = (E, \boldsymbol{p})$	momentum de cuatro dimensiones. 17-7
S	número de “extrañeza”. 2-10
$\boldsymbol{\nabla}$	operador nabla, $\boldsymbol{\nabla} = \boldsymbol{\nabla} = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$. 14-10
W	peso. 4-5
$\boldsymbol{x}_\mu = (t, \boldsymbol{r})$	posición de cuatro dimensiones. 34-10
P	potencia. 24-1
P	presión. 39-5
$P(A)$	probabilidad de observar el evento A. 6-1
$\propto$	proporcional. 5-1
r	radio o distancia. 5-9
$\boldsymbol{r}$	radio (posición) vector. 11-5
γ	relación de capacidad de calor (índice adiabático o relación de calor específico). 39-6
R	resistencia. 23-6
σ	sección transversal. 5-9
T	temperatura absoluta. 39-10
t	tiempo. 5-1
τ	torque. 18-4
W	trabajo. 14-2
i	unidad imaginaria. 22-8
$\vert \quad \vert$	valor absoluto. 6-5
$\langle \quad \rangle$	valor esperado. 6-6
$\boldsymbol{a}$	vector de aceleración. 19-2
$\boldsymbol{E}$	vector de campo eléctrico. 12-11
$\boldsymbol{B}$	vector de campo magnético (inducción magnética). 12-11
$\boldsymbol{L}$	vector de momento angular. 20-4
$\boldsymbol{p}$	vector de momentum. 15-9

$\boldsymbol{k}$	vector de onda. 34-10
$k_\mu = (\omega, \boldsymbol{k})$	vector de onda de cuatro dimensiones. 34-10
$\boldsymbol{\tau}$	vector de torque. 20-5
$\boldsymbol{v}$	vector de velocidad. 11-8
$\boldsymbol{\omega}$	vector de velocidad angular. 20-6
$\boldsymbol{e}_r$	vector unitario en la dirección $\boldsymbol{r}$. 28-2
$\boldsymbol{i}$	vector unitario en la dirección x. 11-11
$\boldsymbol{j}$	vector unitario en la dirección y. 11-11
$\boldsymbol{k}$	vector unitario en la dirección z. 11-11
u	velocidad. 15-2
V	velocidad. 4-6
ω	velocidad angular. 18-8
c	velocidad de la luz, $c = 2.9792458 \times 10^8$ m/s. 4-7
T	vida media. 5-4
V	voltaje. 23-5
V	volumen. 39-3

Los textos contemporáneos de física pueden catalogarse en un antes y después de las lecciones de Richard Feynman, propuestas a partir de las clases impartidas en el Instituto Tecnológico de California (Caltech) de 1961 a 1963. Feynman transformó la concepción de esta pequeña ventana al universo con una nueva propuesta llena de vitalidad en su enfoque. A partir de entonces, la enseñanza de esa ciencia logra profundizar el estudio de los fenómenos físicos de forma puntual y sencilla.

Feynman, Premio Nobel de Física en 1965, dejó en este primer volumen una introducción básica al conocimiento de la física medianta tres ejes fundamentales: la mecánica, la radiación y el calor. En ellos se encierran los temas que parten desde los átomos y la energía, pasan por las leyes de Newton y de la termodinámica, y concluyen en el estudio de las ondas y el sonido. La importancia de la obra —compuesta en tres volúmenes— es tal que continúa formando parte de la bibliografía esencial para quienes estudian física e ingeniería en la actualidad.

La física de Feynman es sobre simplicidad, belleza, unidad y analogía, presentada con entusiasmo y perspicacia.

Nature

Una reacción en cadena no es una mala analogía para la vida de Feynman. A partir de una masa crítica de materia gris se dispara en todas direcciones, produciendo tanto calor como luz.

Time

www.fondodeculturaeconomica.com

FONDO DE CULTURA ECONÓMICA

CONACYT